

238 AAAGTAAGTTTCTGGATTTTGTAGAACTCTGGACGTTGAGGATTCATTATGCTGTGG 297



Db 419 ATGAATGACAAATGAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 478  
QY 497 CTGAAGTATCTCTGAAGAACCCAGCATAGTTACCCAGTGAAGAGACAGCTTTACACAAG 556  
Db 479 ATGAAGAAGACAGATGAAGAACAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 538  
QY 557 ACTGTAATCAGGATATGAAGAGATGATGCTGTGAGCAGGAGAAATCCAGATTCAGTGAAC 616  
Db 539 AACAAGATGAAGAACAAGAAAGTATATGCTGAAAAAGAAATGAAGATGAAGAAAAA 598  
QY 617 CAGTAGTAGAAGATGAAGATGTCACCATGATACAGATGATGTACATACCAAGTCTATG 676  
Db 599 AAGAAAGAAAGAAAGAAAGAAAGATGAAAAATATATGTTGAAAAAGAAAGATGAAG 658  
QY 677 AGGAACAAGCAGTATATGAACTCTGAAATGAAGGATAGAAATCACACAAGTAACTG 736  
Db 659 AAGTATATGTAGAAAAGAAACAAGACATTTTAAAGATTTAGAAAAGATTTTAGAAG 718  
QY 737 CTCCTCCCTGAGGATTAATCTGTAGAAGATTCACAGGTAATTTAGAGAAGATGAAGCATTT 796  
Db 719 AAAAAGTAGAAAGTGAACATAAAGATGATGAAGAGAGACAAAAATTAAGAAGCATGAAG 778  
QY 797 TTCTGTGGAAGAACACAGAGATGATCCACAGATGATTAAGCTTCAAAAAA 850  
Db 779 TTGAAGAAGAACCGAATAAGAAATAAAAGGAGAAATCAGAGAATAGATGAGA 832

RESULT 4  
T28082  
ID T28082 standard; DNA; 86 BP.  
AC T28082;  
DE 31-DEC-1996 (first entry)  
KW Probe 03F1 isolated from fibroblasts.  
KW Polymerase chain reaction; PCR; primer; amplify; human; fibroblast; AIDS;  
KW enhanced differential display; EDD; mRNA preparation; senescent cell;  
KW quiescent cell; dividing cell; senescence-related gene; gene expression;  
KW non-senescent cell; age-related lipofuscin; retina; therapy; liver spot;  
KW donor tissue; senescent melanocyte; melanin; hypopigmentation; ss.  
OS Synthetic.  
PN W09613610-A2.  
PD 09-MAY-1996.  
PR 24-AUG-1995; U11230.  
PR 31-OCT-1994; US-332420.  
PA (GERO-) GERON CORP.  
PI Feng J, Funk W, Hirsch KS, Linskens MHK, Villeponteau B;  
PI West MD;  
DR WPI; 96-251464/25.  
PT Identifying, isolating and regulating senescence-related genes -  
PT useful to ameliorate problems associated with accumulation of  
PT senescent cells, e.g. age-related lipofuscin accumulation in the  
PT retina and AIDS  
PS Claim 8; Page 36; 135pp; English.  
CC T28076-T28113, and T28131-T28173 represent novel senescent-related gene  
CC sequences isolated from fibroblasts using the method of the invention.  
CC In the method of the invention, mRNA is isolated from a senescent cell,  
CC and a young quiescent cell, and the mRNAs are amplified (using primers  
CC such as those shown in T28044-T28075) in separate reaction mixtures. The  
CC amplified sequences are then separated by size or charge, and the  
CC products are analysed to identify a gene from young quiescent cells and  
CC dividing cells, that is present at a different level from senescent  
CC cells. To enhance the method even more, it can be performed in  
CC conjunction with an enhanced differential display (EDD) method (an mRNA  
CC preparation method) on the fibroblasts. The method can be used for the  
CC rapid and efficient identification and isolation of senescence-related  
CC genes and gene products, and to detect and distinguish between senescent  
CC and non-senescent cells. It can also be used to destroy cells expressing  
CC senescence specific (or related) gene products, and to screen for  
CC compounds capable of altering gene expression in senescent cells. The  
CC method can also be used to ameliorate problems associated with the  
CC accumulation of senescent cells such as age-related lipofuscin  
CC accumulation in the retina, and in the treatment of AIDS. Also, the  
CC method can be used to distinguish young cells from senescent cells in

CC donor tissue, which is useful in removing senescent melanocytes  
CC overexpressing melanin which cause hypopigmentation, or liver spots.  
SQ Sequence 86 BP; 33 A; 10 C; 13 G; 30 T;  
Query Match 3.5%; Score 84.4; DB 1; Length 86;  
Best Local Similarity 98.8%; Pred. No. 7.8e-11;  
Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2265 ATAATAAACTCTTCATTTTGTGTAATTATAGAAGCTACTTTTATATAAGCCATATTTT 2324  
Db 1 ATAATAAACTCTTCATTTTGTGTAATTATAGAAGCTACTTTTATATAAGCCATATTTT 60  
QY 2325 TAGGAAACTAAGGAGTGACATAGAA 2350  
Db 61 TAGGAAACTAAGGAGTGACATAGAA 86

RESULT 5  
N60472  
ID N60472 standard; DNA; 4590 BP.  
AC N60472;  
DE 24-AUG-1991 (first entry)  
KW Sequence encoding the ring-infected Erythrocyte Surface Antigen  
DE (RESA).  
KW Malaria vaccine; antigen; epitope; ss.  
OS Plasmodium falciparum.  
FH Key Location/Qualifiers  
FT exon 801..995  
FT /\*tag= a  
FT exon 1199..4225  
FT /\*tag= b  
PN W08601802-A.  
PD 27-MAR-1986.  
PR 11-SEP-1985; 006960.  
PR 11-SEP-1984; AU-007067.  
PR 11-SEP-1984; AU-007066.  
PR 10-SEP-1985; AU-047326.  
PA (HALL-) HALL INST MED RES.  
PI Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;  
DR WPI; 86-094085/14.  
DR P-PSDB; P60569.  
PT DNA coding for Plasmodium falciparum antigens - expressing  
PT poly:peptide(s) having antigenicity of RESA or FIRA antigens of P  
PT falciparum  
PS Claim 4; Fig 1; 55pp; English.  
CC The inventors claim a novel DNA molecule which comprises a  
CC nucleotide sequence corresp. to all or a portion of the base  
CC sequence coding RESA (N60472) or FIRA (N60473). RESA and FIRA have  
CC antigenicity suitable for providing protective immunity against  
CC Plasmodium falciparum malarial infections.  
SQ Sequence 4590 BP; 1933 A; 437 C; 673 G; 1547 T;  
Query Match 2.9%; Score 69.6; DB 1; Length 4590;  
Best Local Similarity 46.0%; Pred. No. 8.2e-07;  
Matches 314; Conservative 0; Mismatches 359; Indels 9; Gaps 2;  
QY 172 GATGGTGATGAGATTTTGTGATGATGCCAAGTTTATTAGGACTTTAAAGAGAGA 231  
Db 3584 CATATTGTAGATATATTATGATGGAATTAGCAATTAATACAGAATGAAAAACCA 3643  
QY 232 TCTATTTCAGCCAGCTCCCGCAGAGAGGCTGAGCCACACACTGAGCCCGAGGAG 291  
Db 3644 AATGAAATGTACCAGACATGTACAACATAATGCTGAAGAAATGTAGAACATGATGCT 3703  
QY 292 CAGGTTCTCTGGAGGAGACACCCAGAAATATCGAAGATGAAGCAAGAAACAAATTCAG 351  
Db 3704 GAAGAAATGTAGACATGATGCTGAAGAAATGTAGAACATGATGCTGAAGAAATGTA 3763  
QY 352 TCCCTTCTCCATGAATGGTACACCCAGAACATGTTGAGGAGAGAGACTTGCACAGAA 411  
Db 3764 GAACATGATGCTGAA---GAAATGTAGAACATGATGCTGAAGAAATGTAGAGAAAT 3820

|    |  |      |  |      |
|----|--|------|--|------|
| QY |  | 412  | GATGGACCCACAGGAAACCACAAGAAGAGTGTATGAGTTCTTATTGGCGACTGATGA        | 471  |
| Dd |  | 3821 | GTTGAAGAAGTAGAAGAAAATGTAGAAGAAAATGTAGAAGAAAATGTGT                | 3880 |
| QY |  | 472  | GATGATAGATTTCAGACCCTGGNA CCT GAAGTGATC TCATGAAGAACCGCACCATAGTTAC | 531  |
| Dd |  | 3881 | GAAGAAGTAGAAGAAAAATGTAGAAGAAAATGTAGAAGAAAATGTGTGAA               | 3940 |
| QY |  | 532  | CACGTGGAAGAGACAGTTTCACAAGACTCTAATCAGGATATGGAAGAGATGCATCTGAG      | 591  |
| Dd |  | 3941 | GAANA GTTGAAGAAAATGTTGAAGAAAATGTAGAAGAAAATGTAGAAGAAAATGTTGAA     | 4000 |
| QY |  | 592  | CAGGAAAATCCAGATTCCAGTGAACCAGTAGTAGAAGATGAAGAATTGCCACCATGATACA    | 651  |
| Dd |  | 4001 | GAATATGATCAACAAAATGTTGAAGAAGTAGAAGAAAATGTAGAAGAAAATGTAGAGAA      | 4060 |
| QY |  | 652  | GATGATGTATA CATACCAAGTCTATGAGGAACAAGCAGTATATGAACCTCTAGA AAAATGAA | 711  |
| Dd |  | 4061 | AATGTTGAAGAAAATGTAGAAGAAAATGTTGAAGAAGTAGAAGAAAATGTAGAAGAAAAT     | 4120 |
| QY |  | 712  | GGGATAGAATCA CAGAAGTA CTGCTCCCC - - - - TGAGGATATCCTGTAGAAGAT    | 765  |
| Dd |  | 4121 | GTAGAAGAAAATGTAGAAGAGAA TGTGAAGAGAA TGTGAAGAGAA TGTGTGAAGAA TAT  | 4180 |
| QY |  | 766  | TCACAGGTAATTGTAGAAGAA GTTAAGCATTTTTCTCTGTGAAGAACACAGCAGGAAGTACCA | 825  |
| Dd |  | 4181 | GATGAAGAAAATGTTCAAGAACACACATGAAGAA TATGATGAAT AAAAAAAAAATATATA   | 4240 |
| QY |  | 826  | CCAGATACTTAAAGCTTCAAAA   | 847  |
| Dd |  | 4241 | TAAAGTTTTAATTTTTATAAA  | 4262 |

|             |  |  |
|-------------|--|--|
| RESULT      | 6  |  |
| Q87588      | Q87588 standard; DNA; 996 BP.  |  |
| Q87588;     | AC   |  |
| 19-DEC-1995 | (first entry)  |  |
| DE          | DNA encoding Leucocytozoan protozoa structural protein epitope.      |  |
| KW          | leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;  |  |
| KW          | leucocytozoanosis; treatment; ss.                                    |  |
| OS          | Leucocytozoan protozoa sp.   |  |
| PN          | J07089995-A.   |  |
| PD          | 04-APR-1995.   |  |
| PF          | 10-SEP-1993; 226078.   |  |
| PR          | 10-SEP-1993; JP-226078.  |  |
| PA          | (DOBU-) DOBUTSUVO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.                    |  |
| PA          | (NISS-) NISSEIKEN KK.  |  |
| DR          | WPI: 95-167252/22.   |  |
| PT          | Immune inducing polypeptide against Leucocytozoan protozoa - useful  |  |
| PT          | in production of vaccines for treatment of leucocytozoanosis in      |  |
| PT          | fowl.  |  |
| PS          | Claim 1; Page 14-15; 20pp; Japanese.                                 |  |
| PS          | Q87587-89 encode polypeptides having a whole or partial epitope of a |  |
| CC          | structural protein of Leucocytozoan protozoa (see #70491-93). The    |  |
| CC          | polypeptides and DNA encoding them are useful in the production of   |  |
| CC          | vaccines for the treatment of leucocytozoanosis of fowl.             |  |
| CC          | Sequence 996 BP; 500 A; 94 C; 221 G; 181 T.                          |  |
| SQ          |  |  |

|        |            |   |     |
|--------|------------|---|-----|
| Qy     | 412        | --GATGGACCCACAGGAGAACCAACAAGAGGATGAATGTTCTTATTGGCGACTGATG           | 469 |
| Db     | 590        | TGGAAGAAGATGTAGTAGAAGAACTAATGTAGTTGAAGAAGTAGTAGAAGAAACACCAG         | 649 |
| Qy     | 470        | TAGATGATAGATTTGACACCTCGAACCTCATGATATCTCATGAAGAAACCGAGCATAGTT        | 529 |
| Db     | 650        | TAGTTTGGAAGAAGAAATGATAGAAGAACTCCTCTAGTTGAAGAAGTAGTAGAAGAAACAC       | 709 |
| Qy     | 530        | ACCACGCTGGAAGACAGACTTTTCACACGACTGTAATCAGGATAATGGAAGAGAGATGATCTG     | 589 |
| Db     | 710        | CAGTAGTTGAAGAAGAGTTGTGTAGAAGAAACACCAAGTAGTGGGAAGAAGAAATGGT---       | 765 |
| Qy     | 590        | AGCAGGAAAATCCAGATTCCAGTGAACACAGTAGTAGAAGATGAAAGATTGCACCATGATA       | 649 |
| Db     | 765        | -AGAAGAAACACCAAGTAGTGGGAAGAGATCTAGTAGAAGAAACACCAAGTAGTTGAAGAAG      | 823 |
| Qy     | 650        | CAGATGATGTACATACCAAGTCTATGAGAGAAACAGCAGTATATGAACCTCTAGAAAATG        | 709 |
| Db     | 824        | AAGTTGTTTGGAAGAACACCAAGTATTTGAAGAAGAAGTTGTAGAAGAAACATCATAGTAGTTG    | 883 |
| Qy     | 710        | AAGGSATGAAGAAATCACAGAAGTAGTCTCCCCTCAGGAGATAATCCTGTAGAGAGATTCAC      | 769 |
| Db     | 884        | AAG---AAGAAATGATAGAAGAAACACCAAGTAGTTGAAGAAAAGTAGTAGAAGAAACAC        | 940 |
| Qy     | 770        | AGGTAATTTGTAGAAGAAAGT 788   |     |
| Db     | 941        | CAGTAGTGGGAAGAAAAAGT 959  |     |
| RESULT | 7          |   |     |
| Q87589 | ID         | Q87589 standard; DNA; 585 BP.                                       |     |
| AC     | AC         | Q87589;   |     |
| DE     | DE         | 19-DEC-1995 (first entry)   |     |
| KE     | DE         | DNA encoding Leucocytozoan protozoa structural protein epitope.     |     |
| KW     | KE         | leucocytozoan protozoa; structural protein; epitope; vaccine; fowl; |     |
| KW     | KW         | leucocytozoanosis; treatment; ss.                                   |     |
| OS     | OS         | Leucocytozoan protozoa sp.  |     |
| PN     | PN         | J07085995-A.  |     |
| PD     | PD         | 04-APR-1995.  |     |
| PF     | PF         | 10-SEP-1993; 226078.  |     |
| PR     | PR         | 10-SEP-1993; JP-226078.   |     |
| PA     | (DOBU-)    | DOBUTSUYO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.                           |     |
| FA     | (NISS-)    | NISSEIKEN KK.   |     |
| DR     | WPI:       | 95-167252/22.   |     |
| PT     | Immun      | Inducing polypeptide against Leucocytozoan protozoa - useful        |     |
| PT     | in pro     | dution of vaccines for treatment of leucocytozoanosis in            |     |
| PT     | fowl.      |   |     |
| PS     | Claim      | 1; Page 16; 20pp; Japanese.   |     |
| CC     | Q87587-89  | encode polypeptides having a whole or partial epitope of a          |     |
| CC     | structural | protein of Leucocytozoan protozoa (see R70491-93). The              |     |
| CC     | polypepti  | des and DNA encoding them are useful in the production of           |     |
| CC     | vaccines   | for the treatment of leucocytozoanosis of fowl.                     |     |
| SQ     | Sequence   | 585 BP; 267 A; 57 C; 155 G; 106 T;                                  |     |

|    | Query Match           | 2.8%  | Score 68.6;        | DB 1;      | Length 585; |
|----|-----------------------|---|--------------------|------------|-------------|
|    | Best Local Similarity | 50.9%;  | Pred. No. 6.4e-07; |            |             |
|    | Matches 254;          | Conservative 0;   | Mismatches 224;    | Indels 21; | Gaps        |
| Qy | 302                   | TGGAGGCGAGACCCACGAATATCGAAGATCAACCAAGAACAATTCAGTCCCTCTCTCC      | 361                |            |             |
| Db | 59                    | TGGTAGAAGAAACACCTGTAGTTGAAGAAGAAATGTAGAAGAAACACCTGTAGTTGAAG     | 118                |            |             |
| Qy | 362                   | ATGAAATGGTACGCGAGACATGTTGAGGAGGAAGACTTGCACCAAGAA                | 412                |            |             |
| Db | 119                   | AAGAAATGATAGAAGAAACACCAAGTAGTGGAGAAGAGATGTCGTCGAAGAAACACCAATAG  | 178                |            |             |
| Qy | 412                   | --GATGACCCACAGGAGAACCAACACAGAGATGATGAGTTTCATATGGCGACTGATG       | 469                |            |             |
| Db | 179                   | TGGAAGAAGATGTAGTAGAAGAACTAATGTAGTTGAAGAAGTAGTAGAAGAAACACCAG     | 238                |            |             |
| Qy | 470                   | TAGATGATAGATTTTGACACCCCTGGAACCTCAAGCTATCTCATGAAGAAACCGGACATAGTT | 529                |            |             |





|   |   |
|---|---|
| DR  | P-PSDB; R98747.   |
| PT  | Antibodies to Plasmodium vivax blood stage antigens - used to               |
| PT  | diagnose malaria and to determine whether P. vivax is the species           |
| PT  | responsible for infection   |
| PS  | Example 4; Column 15-20; 22pp; English.                                     |
| CC  | The present sequence encodes a species-specific Plasmodium vivax            |
| CC  | malarial antigen, PvSP-1. The gene appears to be missing a small            |
| CC  | portion of its 5' end. This protein is secreted into the plasma of          |
| CC  | a susceptible mammalian host after infection. Monoclonal/polyclonal         |
| CC  | antibodies can be utilised in assays used to diagnose malaria, as well      |
| CC  | as to determine whether P. vivax is the species responsible for the         |
| CC  | infection.  |
| SQ  | Sequence 3337 BP; 1304 A; 467 C; 875 G; 691 T;                              |
| <br>  |   |
| Query Match 2.5%; Score 61; DB 1; Length 3337;              |   |
| Best Local Similarity 44.2%; Pred.No. 6.6e-05;              |   |
| Matches 299; Conservative 0; Mismatches 375; Indels 3; Gaps |   |
| <br>  |   |
| QY  | 146 TAGGAAACTAGGAATCTATGATGTCTGGTGGTAGAGATTTTGATGTGGATGATGCCA 205           |
| Db  |   |
| Db  | 1723 TATTTAACAAATCCTCAAGGATCTGTGATGCAGGAGCGCTGTACAGTACCATCAAAG 1782         |
| QY  | 206 AAGTTTTATTAGGACTTTAAGAGAGATCTACTTTCAGAGCCAGCAGCTCCCGCCAGAAGAGG 265      |
| Db  |   |
| Db  | 1783 AAGCACCTGTACAAGTACCAAGTGGCAGTAGGCGCCGCAAGAAGTGCACGAAGGAAT 1842         |
| QY  | 266 CTGAGCCACACACTGAGCCCGAGGACGAGTTCCTGTGGAGCCAGACCCCAGATAATCG 325          |
| Db  |   |
| Db  | 1843 TGATGCACTCCAAGAGACGATTTTCGAATTAGAAGGAAGTGCAGAGCTCCAGAGGAAG 1902        |
| QY  | 326 AAGATGAAGCAAAAAGAACAAATTCAGTCCCTTCTCCATGAAATGGTACACGCAGAAATG 385        |
| Db  |   |
| Db  | 1903 GAGAAATTAGTTTAGAAGGAGAGGAGNACCAACGGAAGAGAGCCAGAGAAGGAGCG 1962          |
| QY  | 386 TTGAGGAGGAAGACTTTC AACAGAGAGATGGACCCAGAGGAGAACCAACAGAGAGGATG 445        |
| Db  |   |
| Db  | 1963 CAACAGAAGGAGAAGTGCCACAGAAGAAGAAATTAGAGGCCAACTCCAGAGGACGATTTGCAAT 2022  |
| QY  | 446 ATGAGTTCTTATGGCGATGTAGTGTAGATGATAGATTTTGAGACCTTGGACCTGGAACCTGAAGTAT 505 |
| Db  |   |
| Db  | 2023 TAGAAGAACCAAGGAGAGAAGATAGAGAAGAACCGTAGAGGGGGAAGAACTGCAGAAG 2082        |
| QY  | 506 CTCATGAAGAAAAACCGAGCATAGTTACCCGTGGAAAGACAGATTTCAAGAGCTGTAATC 565        |
| Db  |   |
| Db  | 2083 GAGAAGAGTGGAAAGGTAACCTGCAGAAGTAGAGA ---AGTGAAGAGGTACTCTCGAC 2139       |
| QY  | 566 AGGATATGGAAGAGATGATGCTGTGAGCAGGAAAAATCCAGATTCAGATTCAGAGTAGTAG 625       |
| Db  |   |
| Db  | 2140 AAGTGAAGAAGTGGAAAGAGTACCAAGAAGAAGTAGAAGAGGTACCCCGCAGAAGTAGAAG 2199     |
| QY  | 626 AAGATGAAGATTGCAACCATGATACAGATGATGTAAACATACCAAGTCTATGAGGAACAAG 685       |
| Db  |   |
| Db  | 2200 AAGTGGAAAGGTAACCAAGAAGTGGAAAGGTTACCAAGAAGAAGTGGGAAGAGTACCAAG 2259      |
| QY  | 686 CAGTATATGAACCTCTAGAAAAATGAAGGATAGAATACACAGAAGTAACCTCTCCCCCTG 745        |
| Db  |   |
| Db  | 2260 AAGAAGTGGNAGAGGTACCAAGAGAAGTGGAGAAGTGGAGAAGTAGAGAAGTAGAAG 2319         |
| QY  | 746 AGGATAATCCTGTAGAAGATTACAGGTAATTTGTAGAAGAAGTAAGCATTTTTCCTGTGG 805        |
| Db  |   |
| Db  | 2320 TACCAGCGGTAGTAGAAGTAGAAGTACCAAGCGTAGTAGAAGAAGAGGTGCCAAGAAG 2379        |
| QY  | 806 AAGAACAGGGAAGTA 822   |
| Db  |   |
| Db  | 2380 TAGAAGAAGAAGAAGAA 2396   |
| <br>  |   |
| RESULT  | 10  |
| Q67190  |   |
| ID  | Q67190 standard; DNA; 9636 BP.  |
| AC  | Q67190;   |
| DT  | 20-FEB-1995 (first entry)   |

| DE  | P. falciparum transmission blocking target antigen Pf5230 DNA.          |
|---|---|
| KW  | Protozoan; transmission blocking target antigen; Pf5230; malaria;       |
| KW  | vaccine; ss.  |
| OS  | Plasmodium falciparum.  |
| FH  | Location/Qualifiers   |
| Key   | 149. .9553  |
| FT  | /*tag= a  |
| PN  | WO9417187-A.  |
| PD  | 04-AUG-1994.  |
| PF  | 18-JAN-1994; U00547.  |
| PR  | 29-JAN-1993; US-010409.   |
| PA  | (USSH ) US DEPT HEALTH & HUMAN SERVICES.                                |
| PA  | (USSH ) US SEC DEPT HEALTH.   |
| PI  | Kaslow DC, Williamson KC;   |
| DR  | WPI: 94-264101/32.  |
| DR  | P-PSDB; R57474.   |
| PT  | New Plasmodium falciparum transmission blocking target antigen -        |
| PT  | useful in antimalarial vaccines, also related DNA, expression           |
| PT  | vectors and transformed cells   |
| PS  | Claim 2; Page 24; 63pp; English.  |
| CC  | The DNA may be used to express Pf5230 in a host cell. It can also       |
| CC  | be used in vaccines (by incorporation into viral vectors which are      |
| CC  | then used to infect host cells) and oligonucleotides derived from it    |
| CC  | can be used to identify homologous proteins in other spp.               |
| SQ  | Sequence 9636 BP; 4156 A; 974 C; 1422 G; 3084 T;                        |
| Query Match 2.4%; Score 59; DB 1; Length 9636;              |   |
| Best Local Similarity 44.7%; Pred. No. 0.00028;             |   |
| Matches 273; Conservative 0; Mismatches 335; Indels 3; Gaps |   |
| Qy  | 261 AGAGGGTGGCCACACACTGAGCCGAGGAGCAGGTTCTCTGTGGAGCGAAGACCCAGAA 320      |
| Db  |   |
| Qy  | 979 AGATGATGAGGAGACAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 1038       |
| Db  |   |
| Qy  | 321 TATCGAAGATGAAGCAAAAGAACAAATTTCAGTCCCTCTCCATGAAATGGTTACACGAGA 380    |
| Db  |   |
| Qy  | 1039 AGAAGAAGAGGAGGAGGAAGATATGATGATTATCTTTTGAAGAAAGTGGGATGAAC 1098      |
| Db  |   |
| Qy  | 381 ACATGTTGAGGAGAGACATTCACACAGAGAGATGGACCCACAGAGAACCCACACAGA 440       |
| Db  |   |
| Qy  | 1099 AGAAGAACAAATTACAAGAGGAACATCAGAGAAGTAGTGTCTGAAATCTTCAGAAGAAG 1158   |
| Db  |   |
| Qy  | 441 GGATGATGAGTTTCTTATGCGCACTGATGATAGATTTCAGACCCCTGGAACCTGA 500         |
| Db  |   |
| Qy  | 1159 TTTTATGATGAGGATGAAGATCTGTAGACCGGATGGAGATATGATAGAGTTGA 1218         |
| Db  |   |
| Qy  | 501 AGTATCTCTGAAGAACCGGACATAGTTTACCACGTGAAGAGACAGTTTCACAAAGACTG 560     |
| Db  |   |
| Qy  | 1219 CGAATATTAGAGACCAAGATGGTGATCTTATGATAGTACAATAAAAATGAAGATG 1278       |
| Db  |   |
| Qy  | 561 TANTCAGGATATGGAAGAGATGATGCTGTGAGCAGGAAATCCAGATTCCTAGTGAACCA 620     |
| Db  |   |
| Qy  | 1279 AGATGAAGAGGTAGGTGAAGAGGTAGGTGAAGAGGTAGGTGAAGAGGTAGGTGAAGAG 1338    |
| Db  |   |
| Qy  | 621 AGTAGAGATGAAGATTGCACCATGATACAGATGATTAACATACCAAGTCTATGAGGA 680       |
| Db  |   |
| Qy  | 1339 AGGTGAAGAGGTAGGTGAAGAGGTAGGTGAAGAGGTAGGTGAAGAGGTAGGTGAAGA 1398     |
| Db  |   |
| Qy  | 681 A---CAAGCAGTATATGAACCTCTAGAAAATGAAGGGATAGAAAATCACAGAAGTAACTGC 737   |
| Db  |   |
| Qy  | 1399 AGGTGAAGAGGTAGGTGAAGGGGTAGGTGAAGAGGTAGGTGAAGAGAGGTGAAGAGGT 1458    |
| Db  |   |
| Qy  | 738 TCCCCCTGAGGATAATCCTGTGTAGAAGATTCACAGGTAATTTAGAGAAGTAAGCATTTT 797    |
| Db  |   |
| Qy  | 1459 AGGTGAAGAGAGGTGAATATGTAGATGAAAAGAAAGCAAGGTGAAATATATATCCATT 1518    |
| Db  |   |
| Qy  | 798 TCTGTGGRAGACACCGAGAGTACCACGATACTTAAAGCTTCAAAAAGACTGCCCC 857         |
| Db  |   |
| Qy  | 1519 TGGTGTATGAAGAGAAAAGATGAAGGTGGAGAAAGTTTTTACCTTATGAAAAGAGCGAGGT 1578 |
| Db  |   |
| Qy  | 858 TACCACCACAG 868   |
| Db  |   |

|        |         |
|--------|---------|
| RESULT | 10      |
| Q67190 |         |
| ID     | Q6719   |
| AC     | Q6719   |
| DT     | 20- FEB |



DE Human p160 cDNA 160.2.  
 KW p160; p62; cytoplasmic; T cell; B cell; development; activation;  
 KW modulation; cellular response; cell proliferation; autoimmune disease;  
 KW p56-lck; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 439..3156  
 FT /tag= a  
 FT /product= p160  
 FT /note= "160.2"  
 PN WO9722255-A1.  
 PD 26-JUN-1997.  
 PF 11-DEC-1996; U19944.  
 PR 19-DEC-1995; US-574959.  
 PA (DAND ) DANA FARBER CANCER INST INC.  
 PI Jung I, Shin J, Strominger JL, Vadlamudi RK;  
 DR WPI; 97-341351/31.  
 DR P-PSDB; W31186.  
 PT cDNA encoding p62 and p160 and corresponding proteins - used in the  
 PT treatment of autoimmune disease and for T and B cell proliferation,  
 PT e.g. for treatment of tumours  
 PS Claim 82; Fig 10; 175pp; English.  
 CC This cDNA sequence encodes a novel p160 (160.2) which is capable  
 CC of activating transcription of a variety of genes upon activation of p62  
 CC and is capable of binding to the p62/p56lck complex to modulate Lck  
 CC function in a manner similar to p62. The genes transcribed in response to  
 CC p160 activation likely include those of which are involved in T or B cell  
 CC development/differentiation, T or B cell activation or production of T or  
 CC B cell specific factors e.g. lymphokines or antibodies. This p160  
 CC polypeptide is also a substrate for serine/threonine kinase activity.  
 CC p160 polypeptides can modulate degradation of cellular proteins e.g. cell  
 CC cycle regulatory proteins stimulating expression of cell cycle dependent  
 CC kinase inhibitors and arresting cell cycle progression at specific  
 CC boundaries to thereby modulate cell proliferation. As p160 boosts B cell  
 CC response it may be used to treat disorders where this is beneficial, e.g.  
 CC infections by pathogenic microorganisms. p160 can be used to expand T  
 CC cell populations for treating infectious diseases or cancer and p160  
 CC inhibitors could reduce B or T cell responses and may be used to treat a  
 CC variety of autoimmune diseases, e.g. diabetes mellitus, arthritis,  
 CC multiple sclerosis allergic reactions, Crohn's diseases etc.  
 SQ Sequence 3211 BP; 649 A; 990 C; 948 G; 624 T;

Query Match 2.2%; Score 53.4; DB 1; Length 3211;  
 Best Local Similarity 47.7%; Pred. No. 0.0035;  
 Matches 156; Conservative 0; Mismatches 171; Indels 0; Gaps 0;  
 QY 271 CCACACACTGAGCCCGAGGAGCAGGTTCTCTGTGGAGGAGACACCCAGCAATATCGAAGAT 330  
 DB 2305 CCACCCAGTTGGTCCCTGAGGGGACTCTGTGGGGAGGAGCCCCCAGCCCTCGGAAGAG 2364  
 QY 331 GAAGCAAAAGACAAATTCAGTCCCTTCTCCATGAATGGTACAGCGCAGCAATGTTGAG 390  
 DB 2365 GATTTGACAGTTATTATATCAACAGCAGTGATGAAGAGGAGGAGGAAGAGAGAG 2424  
 QY 391 GGAGAAAGCTTGCACACAGAGATGGACCCACGAGGAGACCAACACAGAGGATGATGAG 450  
 DB 2425 GAAGAAG 2484  
 QY 451 TTTCTTATGGGAGCTGATGTAGATCATAGATTGAGACCCCTGGAACCTGAAGTATCTCAT 510  
 DB 2485 GAAGACTTTGAGGAAGAGAGAGGATGAAGAGGAATATTTTGAAGAGAGAGAGAGAGAG 2544  
 QY 511 GAAGAAACCGAGCATAGTTACCACTGGAAGAGACAGATTTCACAGACTGTTAATCAGGAT 570  
 DB 2545 GAAGAGAGTTTGGAGAGAGATTTGAGGAGAGAGAGAGGTTAGTGAAGAGAGAGAGAA 2604  
 QY 571 ATGGAAGAGATGATCTCTGAGCAGGAA 597  
 DB 2605 GAGGAGGATGAGGAGGAGAGAGAGAA 2631  
 RESULT 14

Q21769  
 ID Q21769 standard; DNA; 616 BP.  
 AC Q21769.  
 DT 23-JUL-1992 (first entry)  
 DE Sequence encoding Plasmodium falciparum immunogenic peptide IIIC.  
 KW Immunogen; vaccine; malaria; epitope; probe; antibody; ss.  
 OS Plasmodium falciparum.  
 FH Key Location/Qualifiers  
 FT CDS 2..616  
 FT /tag= a  
 PN WO9203552-A.  
 PD 05-MAR-1992.  
 PF 14-AUG-1991; F00667.  
 PR 14-AUG-1990; FR-010363.  
 PA (INSP ) INST PASTEUR.  
 PI Muller-Hill B, Kun J, Schreiber M, Gysin J, Pereira Da Silva L,  
 PI Breton C;  
 DR WPI; 92-096896/12.  
 DR P-PSDB; R21614.  
 PT New polypeptide(s) inducing protective antibodies - having  
 PT mediator function on infected red corpuscles and useful as  
 PT diagnostic agents and vaccines  
 PS Claim 6; Fig 3; 66pp; French.  
 CC The polypeptides of the invention contain at least one sequence  
 CC having at least one epitope characteristic of a protein present on  
 CC the surface of red blood cells (RBC) infected with Plasmodium  
 CC falciparum (P.f.). They are useful as immunogens in protective  
 CC vaccines against malaria, and for in vitro immuno-detection of P.f.  
 CC in tissues or biological fluids. Nucleic acid sequences encoding  
 CC the polypeptides are also claimed, and so are nucleotide probes  
 CC contg. all or part of the nucleic acid sequences.  
 SQ Sequence 616 BP; 272 A; 46 C; 154 G; 144 T;  
 Query Match 2.0%; Score 49.2; DB 1; Length 616;  
 Best Local Similarity 47.0%; Pred. No. 0.017;  
 Matches 187; Conservative 0; Mismatches 208; Indels 3; Gaps 1;  
 QY 339 AGACAAATTCAGTCCCTTCTCCATGAATGGTACAGCAGACATGTTGAGGAGAGA 398  
 DB 184 AGGACAACAAGAAATCTGTTACTGAAGAAATAGTAGTGAGAGGATCATTTACTCAAGA 243  
 QY 399 CTTGCAACAAGAAGATGGACCCACAGAGGAGAACCAACAGAGGATGATGTTCTTAT 458  
 DB 244 CATAGTAGAGGAGAGAGATCATGTTACGGAAGAAATAGTAGTGATGAAGATCGGTTAC 303  
 QY 459 GCGGACTGATGATGATAGATTTGAGACCCCTGGAACCTGAGTATCTCATGAACAAAC 518  
 DB 304 TAAGGAATAGTAGAGGATGAAGAAATAGTTACTGAAGAAATAGTAGAGGATGAAGATC 363  
 QY 519 CGAGCATAGTTTACCACGTGGAGAGACAGTTTTCACAAGACTGTAATCAGGATATGAAGA 578  
 DB 364 TTTTACTGAGAGGTTATAGAGAGAGATCATTAATTGAGAGGTTGAAGATAAGAACCA 423  
 QY 579 GATGATGCTGTAGCAGGAGAAATCCAGATTCAGTGAACAGTAGTAGAAGATGAAGATT 638  
 DB 424 GTTGCTTGAAGAAAGAGAGGATCAGTT---ATTAAAGAAATTAATTGACGAAAAATCACT 480  
 QY 639 GCACCATGATACAGATGATGAACATACCAAGTCTATGAGGACAGCAGTATATGAACC 698  
 DB 481 TACTGAAAAAATAGTTTGAAGAGAGAAATCAGTTACTGAAGAGTTGAAGAAAAAGATTC 540  
 QY 699 TCTAGAAATAGAGGATAGAAATCAGAGAACTAAGT 736  
 DB 541 GGTTAAGGAGAGATGAGAGAGAGAGATAGTTGTTG 578  
 RESULT 15  
 ID N90127  
 AC N90127 standard; DNA; 839 BP.  
 DT 1-NOV-1989 (first entry)  
 DE DNA portion of Plasmodium falciparum rhoptry antigen Ag512



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 1999, 23:10:54 ; Search time 489.51 seconds  
(without alignments)  
7609.827 Million cell updates/sec

Title: US-09-040-485-1

Perfect score: 2442

Sequence: 1 CGGAGCTTGAAGGACACAA.....GGTTAACTTTAAATATTTTT 2442

Scoring table: IDENTITY\_NUC

Searched: 2002476 seqs, 76271212 residues

Database : EST.\*

- 1: em\_est1.\*
- 2: em\_est2.\*
- 3: em\_est3.\*
- 4: em\_est4.\*
- 5: em\_est5.\*
- 6: em\_est6.\*
- 7: em\_est7.\*
- 8: em\_est8.\*
- 9: em\_est9.\*
- 10: gb\_est1.\*
- 11: gb\_est2.\*
- 12: gb\_est3.\*
- 13: gb\_est4.\*
- 14: gb\_est5.\*
- 15: gb\_est6.\*
- 16: gb\_est7.\*
- 17: gb\_est8.\*
- 18: gb\_est9.\*
- 19: gb\_est10.\*
- 20: gb\_est11.\*
- 21: gb\_est12.\*
- 22: gb\_est13.\*
- 23: gb\_est14.\*
- 24: gb\_est15.\*
- 25: gb\_est16.\*
- 26: gb\_est17.\*
- 27: gb\_est18.\*
- 28: gb\_est19.\*
- 29: gb\_est20.\*
- 30: gb\_est21.\*
- 31: gb\_est22.\*
- 32: em\_est10.\*
- 33: em\_est11.\*
- 34: em\_est12.\*
- 35: em\_est13.\*
- 36: em\_est14.\*
- 37: em\_est15.\*
- 38: em\_est16.\*
- 39: em\_est17.\*
- 40: em\_est18.\*
- 41: em\_est19.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 489   | 20.0        | 851    | 15 | AA203198    |
| 2          | 455.4 | 18.6        | 502    | 12 | N31186      |

|   |    |       |      |     |    |            |
|---|----|-------|------|-----|----|------------|
| c | 3  | 439.4 | 18.0 | 600 | 30 | AI084624   |
|   | 4  | 432.8 | 17.7 | 436 | 24 | AA424475   |
|   | 5  | 385.4 | 15.8 | 424 | 15 | AA037341   |
| c | 6  | 372.8 | 15.3 | 585 | 15 | AA176260   |
|   | 7  | 365.8 | 15.0 | 574 | 31 | AI267311   |
| c | 8  | 342.8 | 14.0 | 598 | 12 | N22713     |
|   | 9  | 340.8 | 14.0 | 468 | 10 | T69872     |
|   | 10 | 340.8 | 13.9 | 540 | 13 | W03046     |
|   | 11 | 335   | 13.7 | 350 | 10 | T07611     |
|   | 12 | 333.4 | 13.7 | 347 | 12 | N26443     |
| c | 13 | 326   | 13.3 | 471 | 22 | AA602346   |
| c | 14 | 325.4 | 13.3 | 452 | 26 | AA788915   |
| c | 15 | 325.4 | 13.3 | 474 | 27 | AA827544   |
|   | 16 | 316   | 12.9 | 575 | 13 | W44451     |
| c | 17 | 312.6 | 12.8 | 482 | 26 | AA767510   |
|   | 18 | 307.4 | 12.6 | 392 | 10 | T48492     |
|   | 19 | 307.4 | 12.6 | 392 | 20 | HUMHBC2999 |
| c | 20 | 296.8 | 12.2 | 598 | 12 | N48349     |
|   | 21 | 294.6 | 12.1 | 404 | 11 | H26237     |
|   | 22 | 294.4 | 12.1 | 315 | 19 | C00261     |
|   | 23 | 293.6 | 12.0 | 549 | 15 | AA176259   |
|   | 24 | 290.8 | 11.9 | 379 | 12 | H81879     |
| c | 25 | 281   | 11.5 | 290 | 22 | AA593699   |
| c | 26 | 279   | 11.4 | 426 | 24 | AA157848   |
| c | 27 | 276.6 | 11.3 | 361 | 25 | AA669808   |
| c | 28 | 270   | 11.1 | 417 | 22 | AA576235   |
|   | 29 | 267   | 10.9 | 289 | 17 | AA304742   |
| c | 30 | 263   | 10.8 | 405 | 25 | AA702708   |
|   | 31 | 253   | 10.4 | 279 | 15 | H96029     |
|   | 32 | 250.4 | 10.3 | 265 | 26 | AA745282   |
| c | 33 | 250.2 | 10.2 | 397 | 25 | AA079718   |
| c | 34 | 250   | 10.2 | 404 | 23 | AI290103   |
| c | 35 | 246.8 | 10.1 | 410 | 21 | AA541348   |
| c | 36 | 245.8 | 10.1 | 290 | 11 | H25536     |
| c | 37 | 243.8 | 10.0 | 446 | 15 | AA159372   |
| c | 38 | 239.4 | 9.8  | 412 | 12 | H99385     |
| c | 39 | 234.6 | 9.6  | 454 | 21 | AA523182   |
| c | 40 | 227.4 | 9.3  | 391 | 12 | N23024     |
| c | 41 | 226.4 | 9.3  | 324 | 24 | AA601990   |
| c | 42 | 215.2 | 8.8  | 387 | 15 | H96030     |
| c | 43 | 188.4 | 7.7  | 330 | 12 | H99337     |
| c | 44 | 169.4 | 6.9  | 285 | 12 | H90076     |
|   | 45 | 157   | 6.4  | 160 | 17 | AA309066   |

ALIGNMENTS

RESULT 1

AA203198

LOCUS

DEFINITION

AA203198

clone 446570 5', mRNA sequence.

AA203198

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Euthera; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

1 (bases 1 to 851)

AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfs, T., Soares, M., Tan, F.,

Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and

Wilson, R.

TITLE

The WashU-Merck EST Project

JOURNAL

Unpublished (1995)

COMMENT

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

24-JAN-1997

EST

INFLS S1

Homo sapiens

CDNA

clone 446570 5', mRNA sequence.

AA203198

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Euthera; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

1 (bases 1 to 851)

AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfs, T., Soares, M., Tan, F.,

Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and

Wilson, R.

TITLE

The WashU-Merck EST Project

JOURNAL

Unpublished (1995)

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108



|||||  
Db 181 AAAATTTCTACATTAGCTTTAGTGTTCAGATTAAACACTTTTGAATTTTGTAGCTTTT 240  
QY 1445 AGCTGATTAATTAGAAAAA-TTAATATTTTCAGTGAAGTTTAAATATTCATTTATTTAT 1503  
Db 241 AGCTGATTAATTAGAAAAA-TTAATATTTTCAGTGAAGTTTAAATATTCATTTATTTAT 300  
QY 1504 TTTTAAATGAGGGGAAAGCTGAAATTCCTTGTAAAGACACAAAGGAAAAAGATGGC 1563  
Db 301 TTTTAAATGAGGGGAAAGCTGAAATTCCTTGTAAAGACACAAAGGAAAAAGATGGC 360  
QY 1564 CTTACTATTATCATGCAAAAATGCTTTTGTGGCACCCTCAGATTAATCATATAATAGCTAT 1623  
Db 361 CCTACTATTATCATGCCAAAATGCTTTTGTGGCACCCTCAGATTAATCATATAATAGCTAT 420  
QY 1624 AGCTCTCTT-CACATTTGTTTAAATTTTGAACAACTGTATAAATTTACTGGTGCATAAC- 1682  
Db 421 AGCTCTCTTCCAGCATTTGTTTAAATTTGAACAACTGTATAAATTTACTGGTGCATAACT 480  
QY 1682 TTAAGATTAATTCGCTCTTGG 1703  
Db 481 TTAAGATTAATTCGCCCTTTG 502  
RESULT 3  
LOCUS AI084624 600 bp mRNA EST 01-OCT-1998  
DEFINITION O278c04.x1 Soares\_senescent\_fibroblasts\_NbHSF Homo sapiens cDNA  
clone IMAGE:1681446 3', mRNA sequence.  
ACCESSION AI084624  
NID G3423047  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 600)  
TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
COMMENT Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 1338 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 444.  
Location/Qualifiers  
1..600  
/organism="Homo sapiens"  
/note="Vector: pT73D (Pharmacia) with a modified  
polylinker V\_TPE: phagemid; Site\_1: Not I; Site\_2: Eco  
RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer [5'  
TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M.Fatima Bonaldo."  
/db\_xref="taxon:9506"  
/clone="IMAGE:1681446"  
/clone\_lib="Soares\_senescent\_fibroblasts\_NbHSF"  
/tissue\_type="senescent fibroblast"  
/lab\_host="DH10B (ampicillin resistant)"  
207 a 106 c 74 g 213 t

Query Match 18.0%; Score 439.4; DB 30; Length 600;  
Best Local Similarity 99.1%; Pred. No. 1.9e-86;  
Matches 453; Conservative 0; Mismatches 1; Indels 3; Gaps 1;  
QY 1986 CTATAAGCAAGCCATAAGAAGTGGTTTGATCGATATATTAGGGTAGCTCTTGTATTTTG 2045  
Db 600 CTATAAGCAAGCCATAAGAAGTGGTTTGATCGATATATTAGGGTAGCTCTTGTATTTTG 541  
QY 2046 TTAACATTAAGATAAGTGAAGTCTTTTCCCTCGCTTTTAGGATTAAATCAAGATACCTT 2105  
Db 540 TTAACATTAAGATAAGTGAAGTCTTTTCCCTCGCTTTTAGGATTAAATCAAGATACCTT 481  
QY 2106 CTATATTTTATCATAATAGATCATAGTCTTATATACAAATGAGTGCCTCGCATGGGTA 2165  
Db 480 CTATATTTTATCATAATAGATCATAGTCTTATATACAAATGAGTGCCTCGCATGGGTA 421  
QY 2166 CTCGATGTGTAATGAACCTGAAATAATAAGATAATAAGAAAAAGCAATAAATTTTCTA 2225  
Db 420 CTCGATGTGTAATGAACCTGGA---AATAATAAGATAATAAGAAAAAGCAATAAATTTTCTA 364  
QY 2226 AAGCTGTGCTGTCGGTGATACAGAGATGATCTCAAAATTAATAAAACTCTTCATTTTG 2285  
Db 363 AAGCTGTGCTGTCGGTGATACAGAGATGATCTCAAAATTAATAAAACTCTTCATTTTG 304  
QY 2286 TGAATATAGAAGCTACTTTTATAAAGCCATATTTTTTAGGGAAACTAAGGAGTGACA 2345  
Db 303 TGAATATAGAAGCTACTTTTATAAAGCCATATTTTTTAGGGAAACTAAGGAGTGACA 244  
QY 2346 TAGAACTGATGAATGAGTAAAGTAAAGTTTGTCTGGATTTTGTAGAACTCTGGACGTTG 2405  
Db 243 TAGAACTGATGAATGAGCAAAAGTAAAGTTTGTCTGGATTTTGTAGAACTCTGGACGTTG 184  
QY 2406 AGGATTCATTATGCTGTGGTTAACTTTAAATATATTTT 2442  
Db 183 AGGATTCATTATGCTGTGGTTAACTTTAAATATATTTT 147  
RESULT 4  
LOCUS AA424475  
DEFINITION zv82e07.r1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone  
760164 5', mRNA sequence.  
ACCESSION AA424475  
NID G2103427  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 436)  
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,  
Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,  
White,Y., Wylie,T., Waterston,R. and Wilson,R.  
WashU-Merck EST Project 1997  
Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 425.  
Location/Qualifiers  
1..436  
/organism="Homo sapiens"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from pooled 8-9 week



(total) fetus material with a Not I - oligo(dT) primer [5'  
TGTACCAATCTGAAGTGGAGCGCGCTAAATTTTATTTT 3'].  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo.  
/db\_xref="taxon:9606"  
/clone="760164"  
/clone\_lib="Soares total fetus Nb2HF8 9w"  
/dev\_stage="8-9 weeks"  
/lab\_host="DH10B"  
BASE COUNT 147 a 57 c 70 g 162 t  
ORIGIN

Query Match 17.7%; Score 432.8; DB 24; Length 436;  
Best Local Similarity 99.5%; Pred. No. 5.1e-85;  
Matches 434; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1240 GTTGGACTGCTGTGACACTGTGTTAAACATGATTTAAAGCTATTAAAGTACTTTG 1299  
Db 1 GTTGGACTGCTGTGACACTGTGTTAAACATGATTTAAAGCTATTAAAGTACTTTG 60  
QY 1300 TGTAGCACCTCTTAAAAACGCTAACAGAGATCATCATTTAGCTGTGAAGATTGAGTTGTA 1359  
Db 61 TGTAGCACCTCTTAAAAACGCTAACAGAGATCATCATTTAGCTGTGAAGATTGAGTTGTA 120  
QY 1360 TATACCTGCTGCTGATATCTTATCAAAAATTTTACATTTAGCTTTAAAGTGTTCAGATTAA 1419  
Db 121 TATACCTGCTGCTGATATCTTATCAAAAATTTTACATTTAGCTTTAAAGTGTTCAGATTAA 180  
QY 1420 CACTTTTGAACCTTTGTAGCTTTTACCTGATTAATTAAGAAAATTAATTTTCAAGTAA 1479  
Db 181 CACTTTTGAACCTTTGTAGCTTTTACCTGATTAATTAAGAAAATTAATTTTCAAGTAA 240  
QY 1480 AGTTTAAATTTATCATTTATTTTAAATGAGGGGAAAGCTGAAATTCCTTGT 1539  
Db 241 AGTTTAAATTTATCATTTATTTTAAATGAGGGGAAAGCTGAAATTCCTTGT 300  
QY 1540 TAAGACACAGGAAAAAGAAATGGCCCTACTATTATCATGCAAAAATGCTTTGTTGGCACC 1599  
Db 301 TAAGACACAGGAAAAAGAAATGGCCCTACTATTATCATGCAAAAATGCTTTGTTGGCACC 360  
QY 1600 TCAGATTAATCATATATAGCTATAGTCTCTTCAGCAATTTGTTTAAATTTAGAAAACCT 1659  
Db 361 TCAGATTAATCATATATAGCTATAGTCTCTTCAGCAATTTGTTTAAATTTAGAAAACCT 420  
QY 1660 GTATAAATTACTGGTG 1675  
Db 421 GTATAAATTACTGGTG 436

RESULT 5  
LOCUS AA037341 424 bp mRNA EST 25-NOV-1996  
DEFINITION zc52d12.r1 Soares senescent fibroblasts NBHSF Homo sapiens cDNA  
clone 325943 5', mRNA sequence.  
ACCESSION AA037341  
NID g1512441  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 424)  
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hulman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)

COMMENT

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 815 Std Error: 0.00  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 194.  
Location/Qualifiers  
I. 424  
/organism="Homo sapiens"  
/note="vector: pT73D (Pharmacia) with a modified  
polylinker V-type: phagemid; Site.1: Not I; Site.2: Eco  
RI; TGTACCAATCTGAAGTGGAGCGCGCTAAATTTTATTTT  
3', double-stranded cDNA was size selected, ligated to  
Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of a modified pT73  
vector (Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo."  
/db\_xref="taxon:9606"  
/clone="325943"  
/clone\_lib="Soares senescent fibroblasts NBHSF"  
/tissue\_type="senescent fibroblast"  
/lab\_host="DH10B (ampicillin resistant)"  
BASE COUNT 121 a 80 c 82 g 139 t 2 others  
ORIGIN

FEATURES  
source

Query Match 15.8%; Score 385.4; DB 15; Length 424;  
Best Local Similarity 98.6%; Pred. No. 1.2e-74;  
Matches 419; Conservative 0; Mismatches 3; Indels 3; Gaps 3;  
QY 1712 GTATTCCTCCCTCCAGCACTAGAGACCGCTCAGTCTCTTACTAGTAACTCAGTAAGCG 1771  
Db 1 GTATTCCTCCCTCCAGCACTAGAGACCGCTCAGTCTCTTACTAGTAACTCAGTAAGCG 60  
QY 1772 CTTGAGCTGGTTCATTGAGGATGTGTGAAAGCTCACAGAGCCGATGCTGCTGCTA 1831  
Db 61 CTTGAGCTGGTTCATTGAGGATGTGTGAAAGCTCACAGAGCCGATGCTGCTGCTA 119  
QY 1832 TTTCACGGCAATGAGCTTTTCTTCTACACTGAAGATTTTCTTCTTATTAATGCTGT 1891  
Db 120 TTTCACGGCAATGAGCTTTTCTTCTACACTGAAGATTTTCTTCTTATTAATGCTGT 179  
QY 1892 TTATTTGGGCTCAGAAATAATTCCTGTTGAAAATAATCCTTTGTGAGAAAAGAGGT 1951  
Db 180 TTATTTGGGCTCAGAAATAATTCCTGTTGAAAATAATCCTTTGTGAGAAAAGAGGT 239  
QY 1952 AGCTACACATCATTTTGAAGAGCACTAGCAACTATAAGCAAGCCATGAAGAGTGGT 2011  
Db 240 AGCTACACATCATTTTGAAGAGCACTAGCAACTATAAGCAAGCCATGAAGAGTGGT 299  
QY 2012 TTGATCATATATTAGGGTAGCTCTTGTATTTTGTACATTAAGATGAAGTGAAGTGT 2071  
Db 300 TTGATCATATATTAGGGTAGCTCTTGTATTTTGTACATTAAGATGAAGTGAAGTGT 359  
QY 2072 CCCCTCTTTTAGGATT-AAAATCAAGATACCT-CTATATTTTATCATATAGATCA 2129  
Db 360 CCCCTCTTTTAGGATT-AAAATCAAGATACCT-CTATATTTTATCATATAGATCA 419  
QY 2130 TAGTT 2134  
Db 420 TAGTT 424

RESULT 6

```

AA176260/c  AA176260  585 bp  mRNA  EST  30-DEC-1996
LOCUS      zp29h01.s1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA
DEFINITION clone 610897 3', mRNA sequence.

ACCESSION  AA176260
NID        91757409
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 585)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
TITLE     WashU-Merck EST Project
JOURNAL   Unpublished (1995)
COMMENT   Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@wustl.wustl.edu
            This clone is available royalty-free through LLNL: contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 361.
            Location/Qualifiers
                1..585
                /organism="Homo sapiens"
                /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                XhoI; Cloned unidirectionally. Primer: Oligo dt. NT2
                cells (Ntera-2/c1.D1) induced with Retinoic Acid for 24
                hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5'
                adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor
                sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3"
                /db_xref="taxon:9606"
                /clone_lib="610897"
                /clone_lib="Stratagene neuroepithelium (#937231)"
                /dev_stage="Ntera-2/RA neuroepithelial cells"
                /lab_host="SOLR (kanamycin resistant)"
                complement(<1..>585)
BASE COUNT  193 a 112 c 83 g 189 t 8 others
ORIGIN
mRNA
Query Match 15.3%; Score 372.8; DB 15; Length 585;
Best Local Similarity 91.6%; Pred. No. 6.7e-72;
Matches 413; Conservative 0; Mismatches 34; Indels 4; Gaps 2;

QY 1993 CAAAGCCATAAGAGTGGTTTGATCATATATAGGGGTAGCTTGTGATTTGTTTACAT 2052
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 584 CCAANCCATAGGAAGNGTGTACCCCAATATTGGGGGTAGCCCTCGATTTGTTTACAT 525
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2053 TAAGTAAGGTGACTTTTCCCTCTCTTTAGGATTAATAATCAAGATACTTCTATATT 2112
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 524 TAGGATAGGGGGCTTTTCCCTCTCTTTNGGATNAANAANCAAGGTACTTCTATATT 465
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2113 TTATCACTATAGATCATAGTATTATATACATAGTAGTACTGCTG-CATGGGTACTCGAT 2171
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 464 TTANCACTATAGATCATAGTATTATATACATAGTAGTACTGCTGCTGCGTACTCGAT 405
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2172 GTCATGAACTGAATTAATAAGTAATAAGAAAGCAATATTTCTAAAGCTG 2231
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 404 GTCATGAACTGAATTAATAAGTAATAAGAAAGCAATATTTCTAAAGCTG 348
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2232 TGCTGTGGGTGATACAGATGATGATCTCAATATATATAAATCTCTTCATTTTGTGAATT 2291
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 347 TGCTGTGGGTGATACAGATGATGATCTCAATATATATAAATCTCTTCATTTTGTGAATT 288
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

BASE COUNT 164 a 150 c 109 g 150 t 1 others  
ORIGIN

[illegible]

TITLE WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 338  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

## FEATURES

Source  
1. 468  
/organism="Homo sapiens"  
/clone="81068"  
BASE COUNT 159 a 90 c 114 g 102 t 3 others  
ORIGIN

Query Match 14.0%; Score 340.8; DB 10; Length 468;  
Best Local Similarity 91.6%; Pred. No. 6.5e-65;  
Matches 416; Conservative 0; Mismatches 29; Indels 9; Gaps 5;

```
QY 472 GATGATAGATTGAGACCCCTGGACCTGAAGTATCTCATGAAGAAACCGGACATAGTTAC 531
|||
DB 6 GAGGATAGATTGAGACCCCTGGACCTGAAGTATCTCATGAAGAAACCGGACATAGTTAC 65
|||
QY 532 CAGCTGGAAGACAGAGTTTCAACAAGACTGTAATCAGGATATGGAAGAGATGCTCTGAG 591
|||
DB 66 CAGCTGGAAGACAGAGTTTCAACAAGACTGTAATCAGGATATGGAAGAGATGCTCTGAG 125
|||
QY 592 CAGGAAATCCAGATTCAGTGAACCACTAGTAGAAGATGAAAGATTCACCATGATACA 651
|||
DB 126 CAGGAAATCCAGATTCAGTGAACCACTAGTAGAAGATGAAAGATTCACCATGATACA 185
|||
QY 652 GATGATGATACATACCAAGTCTATGAGAACCAAGCAGTATATGAACCTCTAGAAATGAA 711
|||
DB 186 GATGATGATACATACCAAGTCTATGAGAACCAAGCAGTATATGAACCTCTAGAAATGAA 245
|||
QY 712 GGGATACAAATCACAGAGTAACCTGCTCCCC- TGAGGATATCTGTAGAGATTCACA 770
|||
DB 246 GGGATACAAATCACAGAGTAACCTGCTCCCCCTTGAGGATATCTGTAGAGATTCACA 305
|||
QY 771 GGTAAATGTA- GAAGAGTAGTAGCATTTTCTCTGT- GGAAGAACACAGCAGGAAGTACCACCA 828
|||
DB 306 GGTAAATGTAAGAGTAGTAGCATTTTCTCTGTGGGAGAACAGCAGGAAGTACCACCA 365
|||
QY 829 GATACCTT- AAAGCTTCAAAAGACTGCC- CTACCACACAGGAGGACCGCTTAA 882
|||
DB 366 GNTACTTTAAAGCTTCCAAAAGAGTTGCCCTACCACACAGGNGGGACCGCTTAA 425
|||
QY 883 CCATACCTCCAAAGATGGCTGTGATAGATCTT 916
|||
DB 426 CCTACGGTTCCAAAGGTGGCTGTGTGTGCTT 459
|||
```

## RESULT 10

W03046 540 bp mRNA EST 18-APR-1996  
LOCUS za04d08.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone 291567  
DEFINITION 5', mRNA sequence.  
ACCESSION W03046  
NID 91275093  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 540)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)

TITLE  
JOURNAL  
COMMENT

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: mob.REGA+ET  
High quality sequence stop: 215.

## FEATURES

## Source

1. 540  
/organism="Homo sapiens"  
/note="Vector: pT73D (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGAGCGCGCAGCTTTTCTTTTCTTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library constructed by Bento Soares and  
M. Fatima Bonaldo. RNA from normal foreskin melanocytes  
(FS374) was kindly provided by Dr. Anthony P. Albino."  
/db\_xref="taxon:9606"  
/clone="291567"  
/clone\_lib="Soares melanocyte 2NbHM"  
/sex="Male"  
/lab\_host="DH10B (ampicillin resistant)"  
BASE COUNT 157 a 93 c 124 g 163 t 3 others  
ORIGIN

Query Match 13.9%; Score 340; DB 13; Length 540;  
Best Local Similarity 91.3%; Pred. No. 9.9e-65;  
Matches 439; Conservative 0; Mismatches 28; Indels 14; Gaps 7;

```
QY 1509 TAAATGAGAGGGGAAGCTGAAATTCCTTGTAAACACACAGAAAAGAAATGGCCCTAC 1568
|||
DB 1 TAAATGAGAGGGGAAGCTGAAATTCCTTGTAAACACACAGAAAAGAAATGGCCCTAC 60
|||
QY 1569 TATTATCATGCAAAAATGCTTTGTTGGCACCCTCAGATTAAATCATATATAGCTATAGTCT 1628
|||
DB 61 TATTATCATGCAAAAATGCTTTGTTGGCACCCTCAGATTAAATCATATATAGCTATAGTCT 120
|||
QY 1629 CTTGAGCATTTGTTTAAATTTTAGAAAACCTGTATAAATTACTGGTGCATACCTTAAAGA 1688
|||
DB 121 CTTGAGCATTTGTTTAAATTTTAGAAAACCTGTATAAATTACTGGTGCATACCTTAAAGA 180
|||
QY 1689 TTATTCTGCTTTGGCTAATTGAGTAATTCCTCCCTCCAGCAGTAGACCGCTCAGTGCTC 1748
|||
DB 181 TTATTCTGCTTTGG- TAATTGAGTAATTCCTCCCTCCAGCAGTAGACCGCTCAGTGCTC 239
|||
QY 1749 TTACTAGATGAACCTCAGTAACGCTTTGAGCTGGGTTGATTGAGGATGTGTGAAAAAGCTC 1808
|||
DB 240 TTACTAGATGAACCTCAGTAACGCTTTGAGCTGGGTTGATTGAGGATGTGTG- AAAAGCTC 298
|||
QY 1809 ACAGAGCCCGATGCTGCTGCTATTTTCAC- GGCAATGAGCCCTTTTCTTCTACACTG- 1866
|||
DB 299 ACAGAGCCCGATGCTGCTGCTGCTATTTTCACGGGCAATGAGCCCTTTTCTTCTACACTGGA 358
|||
QY 1866 AGATTTTCTCTTATTAATGCTGCTTTATTTT- ---GGGCTCAGAAAATAATGCTCTCTG 1921
|||
DB 359 AGATTTTCTCTTATTAATGCTGCTTTATTTTGGGGGCTCAGGAAAATAATGCTCTCTG 418
|||
```



|            |  |  |   |
|------------|--|--|---|
| Db         | 241  | GAGTTGTATACCTGCACCTGATATCTTATCAAAAAATTTCTACATTAGCTTTAAGTGT | 300   |
| Qy         | 1412   | CAGATTACACCTTTTGAACACCTTTGTAGCTTTTACGTCGATTAAT             | 1456  |
|            |  |  |   |
|            |  |  |   |
|            |  |  |   |
| Db         | 301  | CAGATTACACCTTTTGAATTTTGTAGCTTTNAGCTGGGTAAT                 | 345   |
| RESULT 13  |  |  |   |
| AA602346/c |  |  |   |
| LOCUS      |  |  |   |
| DEFINITION | AA602346   | 471 bp   | mRNA  |
|            | nc08906.s1   | NCI_CGAP_A41   | Homo sapiens cDNA clone IMAGE:1114043, mRNA |
| ACCESSION  |  |  | EST   |
| NID        | AA602346   |  |   |
| KEYWORDS   | G2436324   |  |   |
| SOURCE     | EST.   |  |   |
| ORGANISM   | human.   |  |   |
|            | Homo sapiens   |  |   |
|            | Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;     |  |   |
|            | Primates; Catarrhini; Hominoidea; Homo.                            |  |   |
| REFERENCE  | 1 (bases 1 to 471)   |  |   |
| AUTHORS    | NCI-CGAP   |  |   |
| TITLE      | http://www.ncbi.nlm.nih.gov/ncicgap.                               |  |   |
|            | National Cancer Institute, Cancer Genome Anatomy Project (CGAP),   |  |   |
|            | Tumor Gene Index   |  |   |
| JOURNAL    | Unpublished (1997)   |  |   |
| COMMENT    |  |  |   |
|            | Contact: Robert Strausberg, Ph.D.                                  |  |   |
|            | Tel.: (301) 496-1550   |  |   |
|            | Email: Robert_Strausberg@nih.gov                                   |  |   |
|            | Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.          |  |   |
|            | Emmert-Buck, M.D., Ph.D.   |  |   |
|            | cDNA Library Preparation: Stratagene, Inc., David B. Krizman,      |  |   |
|            | Ph.D.  |  |   |
|            | cDNA Library Arraying: Greg Lennon, Ph.D.                          |  |   |
|            | DNA Sequencing by: Washington University Genome Sequencing Center  |  |   |
|            | Clone distribution: NCI-CGAP clone distribution information can be |  |   |
|            | found through the I.M.A.G.E. Consortium/LLNL at:                   |  |   |
|            | www-bio.llnl.gov/bbr/image/image.html                              |  |   |

```

Seq primer: -40ml3 fwd. ET from Amersham.
Location/Qualifiers
1. .471
/organism="Homo sapiens"
/note="Organ: adrenal gland; Vector: Bluescript SK-;
Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally.
Primer: Oligo dr. two pooled bulk adrenal adenomas. 5'
adaptor sequence: 5' GAATTCGCACGAG 3' 3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' Average insert
size: 1.6 kb."
/db_xref="taxon:9606"
/clone="IMAGE:1114043"
/clone_lib="NCI_CGAP_AAL"
/tissue_type="adrenal adenoma"
/lab_host="SOLR (kanamycin resistant)"
BASE COUNT      168 a    83 c    56 g    164 t
ORIGIN

Query Match          13.3%; Score 326; DB 22; Length 471;
Best Local Similarity 99.1%; Pred. No. 1.le-61;
Matches 339; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

QY   2101 TACTCTATATTTTATCACAATAGATCATAGTAGTTATTATACAATGTAGTGAGTCTCGCAT 2160
      |||||
Db    471 TACTCTATATTTTATCACAATAGATCATAGTTATTATACAATGTAGTGAGTCTCGCAT 412
      |||||

QY   2161 GGGTACTCGATGTGTAATGAACCTGAAATAAATAAGAATAAAGAAAAGCAATAATT 2220
      |||||
Db    411 GGGTACTCGATGTGTAATGAACCTGA--NATAATAAGATAAATAAGAAAGCAATAATT 355
      |||||

QY   2221 TTCTAAAGCTGTGCTGCGGTGATACAGAGATGATCTCAAATTATATAAAACTCTTCA 2280
      |||||
Db    354 TTCTAAAGCTGTGCTGCGGTGATACAGAGATGATCTCAAATTATATAAAACTCTTCA 295
      |||||

```

|                       |  |   |                         |
|-----------------------|--|---|-------------------------|
| QY                    | 2281   | TTTTGTGAATTTATAGAAGCTACTTTTTTATATAAGCCATATTTTTTTAGGAACTAAGGAG | 2340                    |
| Db                    | 294  | TTTTGTGAATTTATAGAAGCTACTTTTTTATATAAGCCATATTTTTTTAGGAACTAAGGAG | 235                     |
| QY                    | 2341   | TGACATAGAAGCTGATGAATGAGTAAAGTAAAGTTTTCGTGGATTTTTGTAGAACTCTGGA | 2400                    |
| Db                    | 234  | TGACATAGAAGCTGATGAATGAGTAAAGTAAAGTTTTCGTGGATTTTTGTAGAACTCTGGA | 175                     |
| QY                    | 2401   | CGTTGAGGATTCATTATGCTGCTGTTAACTTTAAATATATTTT                   | 2442                    |
| Db                    | 174  | CGTTGAGGATTCATTATGCTGCTGTTAACTTTAAATATATTTT                   | 133                     |
| RESULT                | 14   |   |                         |
| AA788915/c            |  |   |                         |
| LOCUS                 | AA788915   | 452 bp  | EST                     |
| DEFINITION            | ae93104.s1   | Stratagene schizo brain s11                                   | Homo sapiens cDNA clone |
| ACCESSION             | AA788915   |   |                         |
| NID                   | g2849035   |   |                         |
| KEYWORDS              | EST.   |   |                         |
| SOURCE                | human.   |   |                         |
| ORGANISM              | Homo sapiens   |   |                         |
| REFERENCE             | Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.   |   |                         |
| AUTHORS               | 1 (bases 1 to 452)   |   |                         |
|                       | Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wyllie,T., Waterston,R. and Wilson,R.  |   |                         |
| TITLE                 | Washu-NCI human EST Project  |   |                         |
| JOURNAL               | Unpublished (1997)   |   |                         |
| COMMENT               | Contact: Wilson RK<br>Washington University School of Medicine<br>4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108<br>Tel: 314 286 1800<br>Fax: 314 286 1810<br>Email: est@watson.wustl.edu<br>This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.<br>Seq primer: -40m13 fwd. ET from Amersham<br>High quality sequence stop: 451.<br>Location/Qualifiers<br>1. .452<br>/organism="Homo sapiens"<br>/note="Vector: Bluescript SK-; Site_1: EcoRI; Library constructed from S-11 frontal lobe, male, 34 years old, 50% caucasian, 50% Aleutian. Schizophrenic suicide. Random primed into EcoRI site of ZAP II Vector. Mass excised. Custom library. Avg insert length 1.4kb. Material obtained by Johnston N., Torrey, E.F., Yolken R., and the Stanley Neuropathology Consortium - Analysis of RNAs from the Brains of Individuals with Psychiatric Diseases (Unpublished) Stanley Neurovirology Laboratory, Johns Hopkins School of Medicine, Baltimore MD."<br>/db_xref="taxon:9606"<br>/clone="1020535"<br>/clone_lib="Stratagene schizo brain s11"<br>/sex="male"<br>/tissue_type="schizophrenic brain S-11 frontal lobe"<br>/dev_stage="34 years old"<br>/lab_host="SOLR (kanamycin resistant)" |   |                         |
| BASE COUNT            | 163 a  | 81 c  | 55 g 153 t              |
| ORIGIN                |  |   |                         |
| Query Match           | 13.3%;   | Score 325.4;  | DB 26; Length 452;      |
| Best Local Similarity | 98.8%;   | Pred. No. 1.5e-61;  |                         |
| Matches 33;           | Conservative   | 0; Mismatches   | 1; Indels 3; Gaps 1;    |
| QY                    | 2100   | ATACTTCTATATTTTATCACTATAGATCATAGTATTATTATACATATGATGAGTCCTGGA  | 2159                    |

```

Insert Length: 2219   Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 437.
Location/Qualifiers
1..474
/organism="Homo sapiens"
/note="Organ: colon; Vector: Bluescript SK-; Site:1:
EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:
oligo dr. Pooled colon tumors. 5' adaptor sequence: 5'
GAATTCGGCAGCAG 3' 3' adaptor sequence: 5'
CTCGAGTCTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."
/db_xref="taxon:9606"
/clone="IMAGE:1420141"
/clone_lib="NCI-CCAP_Col2"
/sex="mixed"
/tissue_type="colon tumor"
/lab_host="SOLR (kanamycin resistant)"
168 a 83 C 58 g 165 t
BASE COUNT
ORIGIN

```

Search completed: May 13, 1999, 23:23:07  
Job time: 733 sec

\*\*\*\*\*  
W O R L D  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu May 13 15:27:35 1999; MasPar time 2.53 Seconds  
Tabular output not generated.  
63.976 Million cell updates/sec

Title: >US-09-040-485-6  
Description: (1-10) from US09040485.pep  
Perfect Score: 64  
Sequence: 1 APPEDNPVED 10

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq32  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 14.497; Variance 50.537; scale 0.287

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

| SUMMARIES  |       |             |                        | Pred. No. |          |
|------------|-------|-------------|------------------------|-----------|----------|
| Result No. | Score | Query Match | Description            | ID        |          |
| 1          | 48    | 75.0        | Mouse inhibitor of ap  | W19749    | 9.40e+01 |
| 2          | 45    | 70.3        | Retinoblastoma (RB) p  | R36534    | 1.84e+02 |
| 3          | 45    | 70.3        | Retinoblastoma pp110R  | R71680    | 1.84e+02 |
| 4          | 45    | 70.3        | Recombinant pp110RB p  | R71681    | 1.84e+02 |
| 5          | 45    | 70.3        | Retinoblastoma suscep  | W09411    | 1.84e+02 |
| 6          | 45    | 70.3        | Retinoblastoma tumour  | R74271    | 1.84e+02 |
| 7          | 45    | 70.3        | Cancer suppressing gen | R05305    | 1.84e+02 |
| 8          | 45    | 70.3        | Predicted retinoblast  | R06289    | 1.84e+02 |
| 9          | 45    | 70.3        | RAS associated GAP NF  | R59922    | 1.84e+02 |
| 10         | 45    | 70.3        | RAS associated GAP NF  | R59921    | 1.84e+02 |
| 11         | 45    | 70.3        | Human neurofibromin.   | W13280    | 1.84e+02 |
| 12         | 45    | 70.3        | Nfl gene product.      | R22268    | 1.84e+02 |
| 13         | 44    | 68.8        | Vibrio cholerae strai  | R86553    | 2.30e+02 |
| 14         | 44    | 68.8        | Vibrio cholerae El To  | R86555    | 2.30e+02 |
| 15         | 44    | 68.8        | Truncated form of hu   | R08390    | 2.30e+02 |
| 16         | 44    | 68.8        | Vibrio cholerae poly   | R72856    | 2.30e+02 |
| 17         | 44    | 68.8        | Incomplete form of hu  | R07661    | 2.30e+02 |
| 18         | 44    | 68.8        | Sequence encoded by v  | R08391    | 2.30e+02 |

19 44 68.8 979 2 R08338 Complete form of huma 2.30e+02  
20 43 67.2 510 4 R22365 SIVmac239 gag gene pr 2.87e+02  
21 42 65.6 506 1 P80802 Sequence encoded by g 3.57e+02  
22 42 65.6 521 1 P80807 Sequence of gag prote 3.57e+02  
23 41 64.1 14 18 R98376 Potato debranching en 4.44e+02  
24 41 64.1 14 13 R64987 Amylopectin starch de 4.44e+02  
25 41 64.1 164 4 R22310 TNF polypeptide mutel 4.44e+02  
26 41 64.1 360 13 R71677 Human CSBP2. 4.44e+02  
27 41 64.1 360 13 R71676 Human CSBP1. 4.44e+02  
28 41 64.1 403 23 W18084 Human Aurora-2. 4.44e+02  
29 41 64.1 499 8 R42122 NK-1 cellulase. 4.44e+02  
30 41 64.1 520 17 R92156 3'-Hydroxymethylcephe 4.44e+02  
31 41 64.1 592 2 R05227 Sequence of rat C kin 4.44e+02  
32 41 64.1 823 1 P80972 Sequence encoded by 3 4.44e+02  
33 41 64.1 823 1 P81502 delta-endotoxin again 4.44e+02  
34 41 64.1 1163 14 R85147 CryIC/CryIA(B) chimera 4.44e+02  
35 41 64.1 1176 29 W46737 Amino acid sequence o 4.44e+02  
36 41 64.1 1176 28 W47035 Bacillus thuringiensis 4.44e+02  
37 41 64.1 1189 13 R71463 Crystall protein encod 4.44e+02  
38 41 64.1 1189 3 R10193 Insecticidal crystal 4.44e+02  
39 41 64.1 1189 8 R39757 Delta endotoxin. 4.44e+02  
40 41 64.1 1189 2 R10130 Lepidopteran-active t 4.44e+02  
41 41 64.1 1190 14 R85150 CryIC/CryIA(B) altern 4.44e+02  
42 40 62.5 489 29 W42942 Mouse MDM2. 5.51e+02  
43 40 62.5 489 28 W42972 Amino acid sequence o 5.51e+02  
44 40 62.5 489 28 W42997 Amino acid sequence o 5.51e+02  
45 40 62.5 614 15 R82630 70K autoantigen, part 5.51e+02

ALIGNMENTS

RESULT 1  
ID W19749 standard; Protein; 140 AA.  
AC W19749;  
DT 16-SEP-1997 (first entry)  
DE Mouse inhibitor of apoptosis protein homologue MIHE.  
KW Inhibitor of apoptosis protein; IAP; mammalian IAP homologue; MIHE;  
KW degenerative disease; infectious disease; autoimmune disease;  
KW cancer; therapy; diagnosis.  
OS Mus musculus.  
PN W09723501-A1.  
PD 03-JUL-1997.  
PF 20-DEC-1996; AU0827.  
PR 22-DEC-1995; AU-007275.  
PA (AMRA-) AMRAD OPERATIONS PTY LTD.  
PI Vaux DL;  
DR WPI; 97-350966/32.  
DR N-PSDB: T72714.  
PT Isolated protein homologues of viral inhibitors of apoptosis - used  
PT to modulate apoptosis for treatment of degenerative, infectious or  
PT autoimmune diseases and cancer  
PT Claim 11; Page 71-72; 136pp; English.  
CC MIHE (W19749) is a murine homologue of baculovirus inhibitor of  
CC an isolated nucleic acid (see also T72714) obt'd. by a database  
CC search for sequences homologous to a baculovirus IAP repeat (BIR)  
CC consensus sequence (see also W19744). Unlike IAP, MIHE does not  
CC contain a RING finger domain. IAP homologues (see also W19745-48  
CC and W19750-52) and their derivatives and chemical analogues can be  
CC used in methods for modulating apoptosis in animal cells,  
CC specifically for treatment, by inhibition, of degenerative and  
CC infectious disease or, by promotion, of cancer and autoimmune  
CC disease.  
SQ Sequence 140 AA;

Query Match 75.0%; Score 48; DB 23; Length 140;  
Best Local Similarity 62.5%; Pred. No. 9.40e+01;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 69 pddnpjee 76  
Qy 3 PEDNPVED 10



RESULT 2  
 ID R36534 standard; Protein; 928 AA.  
 AC R36534; 1993 (first entry)  
 DT 24-AUG-1993 (first entry)  
 DE Retinoblastoma (RB) protein.  
 KW RB gene product; p56RB portion; cell cycle progression control;  
 KW combination; therapeutic methods; arrest; tumorigenesis;  
 KW regulation; physiological processes; blood cell prodn.;  
 KW gamete prodn.; ss.  
 OS Homo sapiens SR-40 cell line.  
 PN WO9308267-A.  
 PD 29-APR-1993.  
 PE 16-OCT-1992; U08918.  
 PR 17-OCT-1991; US-778510.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Goodrich DW, Lee EYP, Lee WH, Wang NP;  
 DR WPI; 93-152462/18.  
 DR N-PSDB; Q41545.  
 PT Method of controlling cell cycle progression - uses purified  
 PT retinoblastoma protein or fragment, for use in combination with  
 PT therapeutic methods to arrest tumorigenesis  
 PS Claim 10; Fig 9; 68pp; English.  
 CC The sequence is that of the retinoblastoma (RB) protein which may be  
 CC used as part of a method of controlling cell cycle progression which  
 CC may be used in combination with therapeutic methods to arrest  
 CC tumorigenesis in organisms. The cell cycle can be reversibly  
 CC arrested in a convenient and safe manner. The protein is used in  
 CC a compsn. is relatively inexpensive and readily obtainable, and  
 CC shows little or no toxic effects on healthy cells. It is also  
 CC compatible with other methods and devices for regulating certain  
 CC physiological processes of the body, such as blood cell prodn. and  
 CC gamete prodn. Fragments of the protein are soluble in low concns.  
 CC of glycerol thus enhancing their value in pharmaceutical applns.  
 CC Sequence 928 AA;  
 SQ Sequence 928 AA;  
 Query Match 70.3%; Score 45; DB 7; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 Db 27 pppepdpeqd 36  
 :|||:|:|:|  
 QY 1 APPEDNPVED 10

RESULT 3  
 ID R71680 standard; Protein; 928 AA.  
 AC R71680;  
 DT 16-OCT-1995 (first entry)  
 DE Retinoblastoma p110RB protein.  
 KW Rb110; retinoblastoma; cancer; tumor suppressor protein;  
 KW cell cycle; p110RB.  
 OS Homo sapiens.  
 PN WO9507708-A.  
 PD 23-MAR-1995.  
 PE 13-SEP-1994; U10357.  
 PR 13-SEP-1993; US-121108.  
 PA (CANJ-) CANJI INC.  
 PI Goodrich DW, Johnson D, Lee EYP, Lee W, Shepard HM;  
 DR WPI; 95-131179/17.  
 PT Admin. of a functional retinoblastoma polypeptide or protein -  
 PT used to prevent and inhibit prim. and sec. retinoblastoma-linked  
 PT cancers  
 PS Disclosure; Fig.2; 163pp; English.  
 CC By chromosomal walking from the esterase D gene on chromosome 13,  
 CC the retinoblastoma susceptibility gene (RB) was identified on the  
 CC basis of chromosomal location, homologous deletion and tumor-specific  
 CC alterations in expression. Screening of cDNA libraries yielded clone  
 CC RB-5 incorporating the complete RB cDNA sequence given in Q86398 and  
 CC encoding a 110-140 kDa nuclear phosphoprotein named p110RB (R71680).  
 CC Recombinant p110RB was prepared in E. coli, but with a P2A mutation  
 CC for cloning convenience (R71681).  
 CC Sequence 928 AA;  
 SQ Sequence 928 AA;  
 Query Match 70.3%; Score 45; DB 7; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 Db 27 pppepdpeqd 36  
 :|||:|:|:|  
 QY 1 APPEDNPVED 10

RESULT 4  
 ID R71681 standard; Protein; 928 AA.  
 AC R71681;  
 DT 16-OCT-1995 (first entry)  
 DE Recombinant p110RB protein.  
 KW Rb110; retinoblastoma; cancer; tumor suppressor protein;  
 KW cell cycle; p110RB; Escherichia coli.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_difference 2  
 FT /note= "amino acid at position 2 is Pro in the  
 FT natural sequence"  
 PN WO9507708-A.  
 PD 23-MAR-1995.  
 PE 13-SEP-1994; U10357.  
 PR 13-SEP-1993; US-121108.  
 PA (CANJ-) CANJI INC.  
 PI Goodrich DW, Johnson D, Lee EYP, Lee W, Shepard HM;  
 DR WPI; 95-131179/17.  
 PT Admin. of a functional retinoblastoma polypeptide or protein -  
 PT used to prevent and inhibit prim. and sec. retinoblastoma-linked  
 PT cancers  
 PS Disclosure; Fig.30; 163pp; English.  
 CC By chromosomal walking from the esterase D gene on chromosome 13,  
 CC the retinoblastoma susceptibility gene (RB) was identified on the  
 CC basis of chromosomal location, homologous deletion and tumor-specific  
 CC alterations in expression. Screening of cDNA libraries yielded clone  
 CC RB-5 incorporating the complete RB cDNA sequence given in Q86398 and  
 CC encoding a 110-140 kDa nuclear phosphoprotein named p110RB (R71680).  
 CC Recombinant p110RB was prepared in E. coli, but with a P2A mutation  
 CC for cloning convenience (R71681).  
 CC Sequence 928 AA;  
 SQ Sequence 928 AA;  
 Query Match 70.3%; Score 45; DB 13; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 Db 27 pppepdpeqd 36  
 :|||:|:|:|  
 QY 1 APPEDNPVED 10

RESULT 5  
 ID W09411 standard; Protein; 928 AA.  
 AC W09411;  
 DT 22-JUN-1997 (first entry)  
 DE Retinoblastoma susceptibility phosphoprotein p110RB.  
 KW Retinoblastoma susceptibility gene; p110RB; nuclear phosphoprotein;  
 KW cancer; osteosarcoma; fibrosarcoma; glioblastoma; breast cancer;  
 KW polyclonal antibody; diagnosis.  
 OS Homo sapiens.  
 PN US5578701-A.  
 PD 26-NOV-1996.  
 PE 17-SEP-1987; 098612.  
 PR 17-SEP-1987; US-098612.  
 PR 11-JUL-1990; US-550877.  
 PR 14-JUL-1992; US-914039.  
 PR 17-JUN-1993; US-079207.

CC Use of recombinant p110RB will reduce the need for conventional  
 CC radiotherapy or chemotherapy.  
 SQ Sequence 928 AA;  
 Query Match 70.3%; Score 45; DB 13; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 Db 27 pppepdpeqd 36  
 :|||:|:|:|  
 QY 1 APPEDNPVED 10

RESULT 4  
 ID R71681 standard; Protein; 928 AA.  
 AC R71681;  
 DT 16-OCT-1995 (first entry)  
 DE Recombinant p110RB protein.  
 KW Rb110; retinoblastoma; cancer; tumor suppressor protein;  
 KW cell cycle; p110RB; Escherichia coli.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_difference 2  
 FT /note= "amino acid at position 2 is Pro in the  
 FT natural sequence"  
 PN WO9507708-A.  
 PD 23-MAR-1995.  
 PE 13-SEP-1994; U10357.  
 PR 13-SEP-1993; US-121108.  
 PA (CANJ-) CANJI INC.  
 PI Goodrich DW, Johnson D, Lee EYP, Lee W, Shepard HM;  
 DR WPI; 95-131179/17.  
 PT Admin. of a functional retinoblastoma polypeptide or protein -  
 PT used to prevent and inhibit prim. and sec. retinoblastoma-linked  
 PT cancers  
 PS Disclosure; Fig.30; 163pp; English.  
 CC By chromosomal walking from the esterase D gene on chromosome 13,  
 CC the retinoblastoma susceptibility gene (RB) was identified on the  
 CC basis of chromosomal location, homologous deletion and tumor-specific  
 CC alterations in expression. Screening of cDNA libraries yielded clone  
 CC RB-5 incorporating the complete RB cDNA sequence given in Q86398 and  
 CC encoding a 110-140 kDa nuclear phosphoprotein named p110RB (R71680).  
 CC Recombinant p110RB was prepared in E. coli, but with a P2A mutation  
 CC for cloning convenience (R71681).  
 CC Sequence 928 AA;  
 SQ Sequence 928 AA;  
 Query Match 70.3%; Score 45; DB 13; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 Db 27 pppepdpeqd 36  
 :|||:|:|:|  
 QY 1 APPEDNPVED 10

RESULT 5  
 ID W09411 standard; Protein; 928 AA.  
 AC W09411;  
 DT 22-JUN-1997 (first entry)  
 DE Retinoblastoma susceptibility phosphoprotein p110RB.  
 KW Retinoblastoma susceptibility gene; p110RB; nuclear phosphoprotein;  
 KW cancer; osteosarcoma; fibrosarcoma; glioblastoma; breast cancer;  
 KW polyclonal antibody; diagnosis.  
 OS Homo sapiens.  
 PN US5578701-A.  
 PD 26-NOV-1996.  
 PE 17-SEP-1987; 098612.  
 PR 17-SEP-1987; US-098612.  
 PR 11-JUL-1990; US-550877.  
 PR 14-JUL-1992; US-914039.  
 PR 17-JUN-1993; US-079207.

PR 08-APR-1994; US-225099.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Lee EYP, Lee W;  
 DR WPI; 97-020465/02.  
 PT Retinoblastoma phosphoprotein pRB110-specific polyclonal antibody  
 PT - for diagnosing retinoblastoma and other related tumours or  
 PT susceptibility to them  
 PS Disclosure; Fig 2; 21pp; English.  
 CC A 110-114 kDa phosphoprotein product (W09411) of the retinoblastoma  
 CC susceptibility gene is designated pRB110. Its sequence was deduced  
 CC from a full-length cDNA sequence that included the sequence of a  
 CC clone derived from the retinoblastoma cell line Y79. pRB110 is  
 CC primarily located in the cell nucleus and has DNA binding activity.  
 CC The absence or loss of pRB110 mediates oncogenicity. pRB110-  
 CC specific polyclonal antibodies can be utilised in methods of  
 CC diagnosing hereditary predisposition to retinoblastoma or to other  
 CC diseases controlled by the retinoblastoma gene such as  
 CC osteosarcoma, fibrosarcoma, glioblastoma and breast cancer.  
 CC Provision of pRB110 to an individual through molecular induction  
 CC and gene transplanting may be used as a means of suppressing  
 CC tumorigenesis.  
 SQ Sequence 928 AA;

Query Match 70.3%; Score 45; DB 22; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 pppdpeqd 36  
 :|||:|:|  
 QY 1 APPEDNPVED 10

RESULT 6  
 ID R74271 standard; Protein; 928 AA.

AC R74271;  
 DT 28-NOV-1995 (first entry)  
 DE Retinoblastoma tumour suppressor protein.  
 KW Recombinant; adenovirus; expression vector; TSG; small lung cancer;  
 KW hepatocarcinoma; melanoma; retinoblastoma; sarcoma; sickle cell;  
 KW anaemia; Tay-Sach's disease.  
 OS Homo sapiens.  
 PN W09511984-A.  
 PD 04-MAY-1995.  
 PF 25-OCT-1994; U12235.  
 PR 25-OCT-1993; US-142669.  
 PR 19-MAY-1994; US-246007.  
 PA (CANU-) CANJI INC.  
 PI Gregory RJ, Maneval DC, Wills KN;  
 DR WPI; 95-178876/23.  
 DR N-PSDB: Q90059.  
 PT Adenoviral vector with deletion of viral protein IX contains  
 PT foreign gene - esp. encoding tumour suppressor protein for gene  
 PT therapy of tumours, reduces contamination by wild type adenovirus  
 PS Disclosure; Fig 3; 92pp; English.  
 CC The sequence is that of a retinoblastoma tumour protein. The gene  
 CC encoding this protein may be used in a novel method involving a  
 CC recombinant adenovirus expression vector to treat diseases  
 CC associated with the absence of the TSG or the presence of a mutated  
 CC TSG, e.g. many forms of carcinoma, sickle cell anaemia or Tay-Sach's  
 CC disease.  
 CC See also R74272.  
 SQ Sequence 928 AA;

Query Match 70.3%; Score 45; DB 14; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 pppdpeqd 36  
 :|||:|:|  
 QY 1 APPEDNPVED 10

RESULT 7

ID R05305 standard; Protein; 928 AA.  
 AC R05305;  
 DT 11-OCT-1990 (first entry)  
 DE Cancer suppressing gene (CSG) product.  
 KW Cancer; cancer suppressing gene; CSG; 13q14; retinoblastoma;  
 KW RB; ds.  
 OS Homo sapiens.  
 PN W09005180-A.  
 PD 17-MAY-1990.  
 PF 30-OCT-1989; 004808.  
 PR 31-OCT-1988; US-265829.  
 PA (REGC) Univ of California.  
 PI Lee WH, Huang HJS;  
 DR WPI; 90-178822/23.  
 DR N-PSDB: Q04713.  
 PT Controlling cancer -  
 PT by replacing ineffective cancer suppressing gene with cloned,  
 PT active gene.  
 PS Claim 35; Page 86; 105pp; English.  
 CC Gene is taken from human chromosome 13q14 retinoblastoma (RB) cDNA.  
 CC By installing a working CSG, safe and specific treatment and  
 CC prophylaxis can be given to cancer patients.  
 SQ Sequence 928 AA;

Query Match 70.3%; Score 45; DB 1; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 pppdpeqd 36  
 :|||:|:|  
 QY 1 APPEDNPVED 10

RESULT 8

ID R06289 standard; Protein; 928 AA.  
 AC R06289;  
 DT 13-DEC-1990 (first entry)  
 DE Predicted retinoblastoma gene product.  
 KW Osteosarcoma; fibrosarcoma; glioblastoma; breast cancer; ds.  
 OS Homo sapiens.  
 PN US4942123-A.  
 PD 17-JUL-1990.  
 PF 17-SEP-1987; 098612.  
 PR 17-SEP-1987; US-098612.  
 PA (REGC ) UNIV OF CALIFORNIA.  
 PI Lee WH, Eva Y, Lee HP;  
 DR WPI; 90-245977/32.  
 PT Diagnosing absence or inactivation of retinoblastoma gene - by  
 PT detecting the absence of specific anti-pRB 110 antibody  
 PT immuno-complex formed using tissue  
 PS Disclosure; P; English.  
 CC Labelled Abs raised to the RB gene product may be used to screen  
 CC for RB and in diagnosis of susceptibility to associated secondary  
 CC cancers such as osteosarcoma, fibrosarcoma, glioblastoma and  
 CC breast cancer.  
 SQ Sequence 928 AA;

Query Match 70.3%; Score 45; DB 1; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 pppdpeqd 36  
 :|||:|:|  
 QY 1 APPEDNPVED 10

RESULT 9

ID R59922 standard; Protein; 2485 AA.  
 AC R59922;  
 DT 22-FEB-1995 (first entry)  
 DE RAS associated GAP NF204.  
 KW Ras; GTPase activating protein; GAP; GAP related domain; GRD;  
 KW pK10; pK11; Saccharomyces cerevisiae; RAS2; v-Ras; heat shock;

KW neurofibromatosis type 1; NFI.  
 OS Homo sapiens.  
 PN W09416069-A.  
 PD 21-JUL-1994.  
 PF 12-JAN-1994; U00198.  
 PR 15-JAN-1993; US-004824.  
 PA (SCHE ) SCHERING CORP.  
 PI Kaziro Y, Nakafuku M;  
 DR WPI: 94-249216/30.  
 PT Blocking Ras-induced effects on a cell - by introducing a GTPase  
 PT activating protein to the cell, used esp. in treatment of cancers  
 PS Disclosure; Page 44-52; 87pp; English.  
 CC Human neurofibromatosis type 1 (NFI)-GAP related domain (GRD) was  
 CC cloned into the yeast expression vector pK10 to obtain pKp11. The  
 CC pKp11 DNA was mutagenized by hydroxylamine in vitro and transformed  
 CC into S. cerevisiae TK161-R2V-D, which carries an oncogenic-type  
 CC RAS2Val19 mutation. The heat shock sensitivity of the clones was  
 CC checked. Plasmid DNAs were recovered, re-transformed into TK161-  
 CC R2V-D, and phenotypic reversion was examined. 2 Clones, NF201  
 CC (given in R59921) and NF204, which had strong suppression activity  
 CC for RAS2Val19, were selected. The mutant NFI-GRDs were also able  
 CC to inhibit v-Ras-induced transformation in mammalian cells.  
 SQ Sequence 2485 AA;

Query Match 70.3%; Score 45; DB 11; Length 2485;  
 Best Local Similarity 60.0%; Pred. No. 1.84e+02;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1192 gppehkpvd 1201  
 :||| :|||  
 QY 1 APPEDNPVED 10

RESULT 10  
 ID R59921 standard; protein; 2485 AA.

AC R59921;  
 DT 22-FEB-1995 (first entry)  
 DE RAS associated GAP NF201.  
 KW Ras; GTPase activating protein; GAP; GAP related domain; GRD;  
 KW pK10; pKp11; Saccharomyces cerevisiae; RAS2; v-Ras; heat shock;  
 KW neurofibromatosis type 1; NFI.  
 OS Homo sapiens.  
 PN W09416069-A.  
 PD 21-JUL-1994.  
 PF 12-JAN-1994; U00198.  
 PR 15-JAN-1993; US-004824.  
 PA (SCHE ) SCHERING CORP.  
 PI Kaziro Y, Nakafuku M;  
 DR WPI: 94-249216/30.  
 PT Blocking Ras-induced effects on a cell - by introducing a GTPase  
 PT activating protein to the cell, used esp. in treatment of cancers  
 PS Disclosure; Page 36-44; 87pp; English.  
 CC Human neurofibromatosis type 1 (NFI)-GAP related domain (GRD) was  
 CC cloned into the yeast expression vector pK10 to obtain pKp11. The  
 CC pKp11 DNA was mutagenized by hydroxylamine in vitro and transformed  
 CC into S. cerevisiae TK161-R2V-D, which carries an oncogenic-type  
 CC RAS2Val19 mutation. The heat shock sensitivity of the clones was  
 CC checked. Plasmid DNAs were recovered, re-transformed into TK161-  
 CC R2V-D, and phenotypic reversion was examined. 2 Clones, NF201 and  
 CC NF204 (given in R59922), which had strong suppression activity for  
 CC RAS2Val19, were selected. The mutant NFI-GRDs were also able to  
 CC inhibit v-Ras-induced transformation in mammalian cells.  
 SQ Sequence 2485 AA;

Query Match 70.3%; Score 45; DB 11; Length 2485;  
 Best Local Similarity 60.0%; Pred. No. 1.84e+02;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1192 gppehkpvd 1201  
 :||| :|||  
 QY 1 APPEDNPVED 10

RESULT 11  
 ID W13280 standard; Protein; 2818 AA.  
 AC W13280;  
 DT 05-JUN-1997 (first entry)  
 DE Human neurofibromin.  
 KW Human; neurofibromin; yeast; IRA; protein; inhibition; GTPase;  
 KW regulation; ras-CAMP; pathway; mammalian; GAP; ras p21; gene;  
 KW activation; neurofibromatosis; type 1; NFI; somatic; mutation;  
 KW tumour; detection; diagnosis; prognosis; defective; treatment.  
 OS Homo sapiens.  
 FH Key  
 FT Location/Qualifiers  
 FT 1175..1534  
 FT /note= "GTPase activating protein (GAP) related  
 FT domain (GRD)"  
 FT 1389..1391  
 FT /note= "conserved region in GRD"  
 FT US5605799-A.  
 PN 25-FEB-1997.  
 PD 551531.  
 PF 12-JUL-1990; US-551531.  
 PR 16-APR-1993; US-047088.  
 PR 28-MAR-1995; US-411389.  
 PA (UTAH ) UNIV UTAH RES FOUND.  
 PI Cawthon RM, Li Y, White RL;  
 DR WPI: 97-153572/14.  
 DR N-PSDB; T46941.  
 DT Detection of defective ras regulation at the neurofibromatosis type  
 DT 1 gene in tumour - by detecting mutation in specified region of gene  
 PS Claim 1; Columns 17-38; 35pp; English.  
 CC The present sequence is human neurofibromin (hNF), which is  
 CC largely homologous to yeast IRA protein (inhibitory regulators of  
 CC the ras-CAMP pathway) and mammalian GAP (ras p21 GTPase activating  
 CC proteins). The hNF gene is the human neurofibromatosis type 1 (NFI)  
 CC gene, somatic mutations of which in the region spanning nucleotides  
 CC 3809-4888 of the NFI cDNA, in human tumours, indicates defective  
 CC ras regulation. Therefore a tumour found to contain a somatic  
 CC mutation in the NFI gene can be treated using ras activity as the  
 CC focus, whereas a tumour not containing such a mutation will require  
 CC other courses of treatment. A tumour containing a somatic mutation  
 CC in the NFI gene can be treated by inactivating ras p21, also as GAP  
 CC p120 is present, but apparently latent, inhibition of GDP/GTP exchange would  
 CC be beneficial and finally inhibition of GDP/GTP exchange would also  
 CC counteract the loss of hNF or hNF GAP related domain activity.  
 CC Sequence. 2818 AA;  
 SQ Query Match 70.3%; Score 45; DB 21; Length 2818;  
 Best Local Similarity 60.0%; Pred. No. 1.84e+02;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1526 gppehkpvd 1535  
 :||| :|||  
 QY 1 APPEDNPVED 10

RESULT 12  
 ID R22268 standard; Protein; 2818 AA.  
 AC R22268;  
 DT 06-MAY-1992 (first entry)  
 DE Nf1 gene product.  
 KW von Recklinghausen neurofibromatosis disease; autosomal dominant;  
 KW gene therapy.  
 OS Homo sapiens.  
 PN W09200387-A.  
 PD 09-JAN-1992.  
 PF 28-JUN-1991; U04624.  
 PR 29-JUN-1990; US-547090.  
 PA (UNMI ) UNIV OF MICHIGAN.  
 PI Collins FS, Wallace MR, Marchuk DA, Andersen LB, Gutmann DH;  
 DR WPI: 92-041568/05.  
 DR N-PSDB; Q20602.  
 PT DNA sequences to von-Recklinghausen neurofibromatosis gene - and  
 PT derived amino acid sequences and probes for screening NFI in early  
 PT stages of disease

PS Claim 25; Page 67; 122pp; English.  
 CC This is the amino acid sequence of the von Recklinghausen neuro-  
 CC fibromatosis (Nf1) gene product. It and antibodies raised to it  
 CC can be used in hybridisation and immunological assays to screen for  
 CC the presence of a normal or defective Nf1 gene product. Functional  
 CC assays to measure levels of gene function can also be used for  
 CC diagnosis or to monitor treatment. Patient therapy through  
 CC supplementation with the normal Nf1 product which can be  
 CC produced by recombinant techniques is also possible.  
 SQ Sequence 2818 AA;

Query Match 70.3%; Score 45; DB 4; Length 2818;  
 Best Local Similarity 60.0%; Pred. No. 1.84e+02;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1526 gppchpvpd 1535  
 :|||:||||  
 QY 1 APPEDNPVED 10

RESULT 13  
 ID R86553 standard; Protein; 395 AA.  
 AC R86553;  
 DT 31-MAR-1996 (first entry)  
 DE Vibrio cholerae strain 395 accessory cholera enterotoxin ACE protein.  
 KW Bacterium; enteropathogen; cholera enterotoxin; choleragen;  
 KW cholera toxin; diarrhea; vaccine; plasmid pCV620; ACE.  
 OS Vibrio cholerae.  
 PN US5470729-A.  
 PD 28-NOV-1995.  
 PF 04-MAR-1983; 472276.  
 PR 04-MAR-1983; US-472276.  
 PR 17-FEB-1984; US-581406.  
 PR 27-MAY-1986; US-867633.  
 PR 05-JUN-1989; US-363383.  
 PR 05-JUN-1990; US-533315.  
 PR 16-JAN-1992; US-821872.  
 PR 12-AUG-1992; US-931943.  
 PA (UYMA-) UNIV MARYLAND BALTIMORE.  
 PI Baudry-maurelli B, Fasano A, Kaper JB;  
 DR WPI; 96-019870/02.  
 DR N-PSDB; T06601.  
 DT New avirulent Vibrio cholerae strains - comprise deletions in the  
 DT cholera toxin and zonula occludens toxin genes, for vaccination  
 PT against cholera  
 PT Example 12; Column 29-36; 55pp; English.  
 CC This is the 11 kDa protein sequence of accessory cholera  
 CC enterotoxin encoded by a 297 bp ORF (accessory cholera enterotoxin;  
 CC ACE) in plasmid pCV620. This protein may be useful in the  
 CC construction of cholera vaccines.  
 SQ Sequence 395 AA;

Query Match 68.8%; Score 44; DB 16; Length 395;  
 Best Local Similarity 55.6%; Pred. No. 2.30e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 121 sppsdspvd 129  
 :|||:||||  
 QY 1 APPEDNPVE 9

RESULT 14  
 ID R86555 standard; Protein; 395 AA.  
 AC R86555;  
 DT 31-MAR-1996 (first entry)  
 DE Vibrio cholerae El Tor accessory cholera enterotoxin ACE protein.  
 KW Bacterium; enteropathogen; cholera enterotoxin; choleragen;  
 KW cholera toxin; diarrhea; vaccine; plasmid pCV620; ACE.  
 OS Vibrio cholerae.  
 PN US5470729-A.  
 PD 28-NOV-1995.  
 PF 04-MAR-1983; 472276.  
 PR 04-MAR-1983; US-472276.

PR 17-FEB-1984; US-581406.  
 PR 27-MAY-1986; US-867633.  
 PR 05-JUN-1989; US-363383.  
 PR 05-JUN-1990; US-533315.  
 PR 16-JAN-1992; US-821872.  
 PR 12-AUG-1992; US-931943.  
 PA (UYMA-) UNIV MARYLAND BALTIMORE.  
 PI Baudry-maurelli B, Fasano A, Kaper JB;  
 DR WPI; 96-019870/02.  
 DR N-PSDB; T06602.  
 DT New avirulent Vibrio cholerae strains - comprise deletions in the  
 DT cholera toxin and zonula occludens toxin genes, for vaccination  
 PT against cholera  
 PT Example 12; Column 37-44; 55pp; English.  
 CC This is the 11 kDa protein sequence of accessory cholera  
 CC enterotoxin encoded by a 297 bp ORF (accessory cholera enterotoxin;  
 CC ACE) in plasmid pCV620. This protein may be useful in the  
 CC construction of cholera vaccines.  
 SQ Sequence 395 AA;

Query Match 68.8%; Score 44; DB 16; Length 395;  
 Best Local Similarity 55.6%; Pred. No. 2.30e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 121 sppsdspvd 129  
 :|||:||||  
 QY 1 APPEDNPVE 9

RESULT 15  
 ID R08390 standard; protein; 437 AA.  
 AC R08390;  
 DT 25-FEB-1991 (first entry)  
 DE Truncated form of human FR-X protein.  
 KW MHC class II; down regulation; autoimmune disease; HLA promoter.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT domain 255..376  
 FT W09012812-A.  
 PN 01-NOV-1990.  
 PF 18-APR-1990; E00625.  
 PR 18-APR-1989; EP-106944.  
 PR 14-AUG-1989; EP-115008.  
 PA (MACH/) Mach B.  
 PI Mach B;  
 DT WPI; 90-348429/46.  
 PT Purified proteins and compans. - regulate expression of MHC class II  
 PT genes and bind to controlling DNA sequences.  
 PS Claim 6; Fig 10; 83pp; English.  
 CC The protein which regulates the expression of MHC class II genes by  
 CC binding to DNA sequences which control this expression. It causes  
 CC down-regulation, useful for the prevention and treatment of auto-  
 CC immune diseases such as Insulin Dependent diabetes, Multiple  
 CC Sclerosis, Lupus Erythematosus and Rheumatoid Arthritis. It can  
 CC also be used for screening and identifying substances capable of  
 CC inhibiting the expression of the MHC II genes. A polypeptide  
 CC comprising only the DNA binding domain (see feature table) will  
 CC bind to the X box of the HLA class II promoter.  
 CC See also R08338, R07661 and R08391.  
 SQ Sequence 437 AA;

Query Match 68.8%; Score 44; DB 2; Length 437;  
 Best Local Similarity 55.6%; Pred. No. 2.30e+02;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 53 sppsdspvq 61  
 :|||:||||  
 QY 1 APPEDNPVE 9

Search completed: Thu May 13 15:27:52 1999  
 Job time : 17 secs.



\*\*\*\*\*  
 M P E R L A  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 13 15:29:15 1999; MasPar time 4.04 Seconds  
 123.105 Million cell updates/sec

Tabular output not generated.

Title: >US-09-040-485-6  
 Description: (1-10) from US09040485.pep  
 Perfect Score: 64  
 Sequence: 1 APPENDNPVED 10

Scoring table: PAM 150  
 Gap 15

Searched: 165420 seqs, 49795644 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: sprenbl6  
 1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
 5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
 9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
 13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 21.054; Variance 25.580; scale 0.823

pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description            | Pred. No. |
|------------|-------|-------------|--------|----|------------------------|-----------|
| 1          | 49    | 76.6        | 225    | 11 | APC-BINDING PROTEIN EB | 2.59e+00  |
| 2          | 48    | 75.0        | 140    | 11 | TIAP.                  | 4.27e+00  |
| 3          | 48    | 75.0        | 1175   | 4  | NPAT.                  | 4.27e+00  |
| 4          | 48    | 75.0        | 1427   | 4  | Q14207                 | 4.27e+00  |
| 5          | 47    | 73.4        | 129    | 2  | E14 PROTEIN.           | 6.99e+00  |
| 6          | 46    | 71.9        | 142    | 4  | HYPOTHETICAL 13.0 KD P | 1.14e+01  |
| 7          | 46    | 71.9        | 142    | 4  | APOTOTIS INHIBITOR SU  | 1.14e+01  |
| 8          | 46    | 71.9        | 146    | 10 | GLUTATHIONE S-TRANSFER | 1.14e+01  |
| 9          | 45    | 70.3        | 2042   | 5  | 3D7VAR1 (FRAGMENT).    | 1.14e+01  |
| 10         | 45    | 70.3        | 53     | 4  | RETINOBLASTOMA SUSCEPT | 1.83e+01  |
| 11         | 45    | 70.3        | 86     | 2  | HYPOTHETICAL 9.9 KD PR | 1.83e+01  |
| 12         | 45    | 70.3        | 97     | 5  | DNA FRAGMENT WITH HIGH | 1.83e+01  |
| 13         | 45    | 70.3        | 258    | 2  | DEHYDROGENASE.         | 1.83e+01  |
| 14         | 45    | 70.3        | 287    | 5  | C03H5.3 PROTEIN.       | 1.83e+01  |
| 15         | 45    | 70.3        | 317    | 4  | RNA POLYMERASE III SUB | 1.83e+01  |
| 16         | 45    | 70.3        | 339    | 2  | JUGLANDIS COPPER-RESIS | 1.83e+01  |
| 17         | 45    | 70.3        | 378    | 4  | NEUROFIBROMATOSIS PROT | 1.83e+01  |
| 18         | 45    | 70.3        | 840    | 4  | NEUROFIBROMATOSIS 1 (F | 1.83e+01  |
| 19         | 45    | 70.3        | 1294   | 4  | KIAA0293 (FRAGMENT).   | 1.83e+01  |
| 20         | 45    | 70.3        | 1426   | 11 | CUT-LIKE 2 (CUX-2).    | 1.83e+01  |
|            |       |             | 2172   | 5  | HYPOTHETICAL 236.2 KD  | 1.83e+01  |
|            |       |             | Q09515 |    |                        |           |

21 45 70.3 2820 11 P97526 NEUROFIBROMIN. 1.83e+01  
 22 44 68.8 109 5 O02443 LARVAL CUTICULAR PROTE 2.94e+01  
 23 44 68.8 122 2 O68696 HYPOTHETICAL 13.8 KD P 2.94e+01  
 24 44 68.8 131 1 Q49612 FORMYLMETHANOFURAN DEH 2.94e+01  
 25 44 68.8 259 1 O29756 CONSERVED HYPOTHETICAL 2.94e+01  
 26 44 68.8 273 13 Q91257 LMPX OF LAMPREY (EC 3. 2.94e+01  
 27 44 68.8 304 2 Q60225 BETA-LACTAMASE. 2.94e+01  
 28 44 68.8 334 10 Q24272 ACTIN 7 (FRAGMENT). 2.94e+01  
 29 44 68.8 348 13 Q57378 RECEPTOR-ASSOCIATED PR 2.94e+01  
 30 44 68.8 365 4 O60509 NSP-LIKE 1 (FRAGMENT). 2.94e+01  
 31 44 68.8 473 5 Q26236 BETA-TUBULIN. 2.94e+01  
 32 44 68.8 598 2 Q50866 KEFC. 2.94e+01  
 33 44 68.8 739 2 Q59557 T-CATALASE (EC 1.11.1. 2.94e+01  
 34 44 68.8 740 2 O05763 CATALASE-PEROXIDASE. 2.94e+01  
 35 44 68.8 1262 5 Q20684 F52H3.7. 2.94e+01  
 36 44 68.8 1506 13 P79927 INTEGRATARY MUCIN B. 2.94e+01  
 37 44 68.8 2764 5 O01399 NEUROFIBROMIN. 2.94e+01  
 38 44 68.8 2802 5 O01398 NEUROFIBROMIN. 2.94e+01  
 39 44 68.8 2802 5 O01397 NEUROFIBROMIN. 2.94e+01  
 40 44 68.8 3413 2 O54593 POLYKETIDE SYNTHASE. 2.94e+01  
 41 43 67.2 149 14 P90432 GAG (FRAGMENT). 4.68e+01  
 42 43 67.2 506 14 O57058 GAG POLYPROTEIN. 4.68e+01  
 43 43 67.2 506 14 O41911 GAG POLYPROTEIN. 4.68e+01  
 44 43 67.2 507 14 P89153 GAG PROTEIN. 4.68e+01  
 45 43 67.2 1103 14 O41928 SSDNA BINDING PROTEIN. 4.68e+01

## ALIGNMENTS

RESULT 1  
 ID Q61167 PRELIMINARY; PRT; 225 AA.  
 AC O61167;  
 DT 01-NOV-1996 (TREMREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
 DE APC-BINDING PROTEIN EB2 (FRAGMENT).  
 GN EB2.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA HOFFMAN N.G., KAY B.K.;  
 RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; U51204; G1256436; -;  
 FT NON\_TER 1 1  
 SQ SEQUENCE 225 AA; 25678 MW; 812588B4 CRC32;

Query Match 76.6%; Score 49; DB 11; Length 225;  
 Best Local Similarity 60.0%; Pred. No. 2.59e+00;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 207 APPEDDEIEE 216  
 |||||:|:  
 Qy 1 APPEDNPVED 10

RESULT 2  
 ID O70201 PRELIMINARY; PRT; 140 AA.  
 AC O70201;  
 DT 01-AUG-1998 (TREMREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)  
 DE TIAP.  
 GN TIAP.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA KOBAYASHI K., OTAKI M., OGASAWARA T., TOKUHISA T.;  
 RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; AB013819; D1029206; -  
SQ SEQUENCE 140 AA; 16297 MW; 2BD48871 CRC32;

Query Match 75.0%; Score 48; DB 11; Length 140;  
Best Local Similarity 62.5%; Pred. No. 4.27e+00;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 69 PDDNPIEE 76  
I:||||:;  
Qy 3 PEDNPVED 10

RESULT 3  
ID Q13632 PRELIMINARY; PRT; 1175 AA.

AC Q13632;  
DT 01-NOV-1996 (TREMREL. 01, CREATED)  
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
DE NPAT.  
GN NPAT.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHEN X., YANG L., UDAR N., LIANG T., XU S., UHRHAMMER N., BAY J.O.,  
RA WANG Z., DANDAKAR U., CHIPLONKAR S., KLISAK I., TELATAR M., YANG H.,  
RA CANNON P., GATTI R.A.,  
RA MAMM. GENOME 0:0-0(0).  
RL EMBL; U58852; G1381667; -  
SQ SEQUENCE 1175 AA; 126731 MW; CFC49BD6 CRC32;

Query Match 75.0%; Score 48; DB 4; Length 1175;  
Best Local Similarity 60.0%; Pred. No. 4.27e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 688 TPPECTPVEN 697  
I:||||:;  
Qy 1 APPEDNPVED 10

RESULT 4  
ID Q14207 PRELIMINARY; PRT; 1427 AA.

AC Q14207; Q14967;  
DT 01-NOV-1996 (TREMREL. 01, CREATED)  
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)  
DE E14 PROTEIN.  
GN E14.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA COOPER P.R., BYRD P.J., TAYLOR A.M.R.;  
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA BYRD P.J., MCCONVILLE C.M., COOPER P.R., PARKHILL J., MCGUIRE G.,  
RA STANKOVIC T., THICK J., TAYLOR A.M.R.;  
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA TISSUE-PLACENTA, AND TESTIS;  
RA MEDLINE; 96338579.  
RA IMAI T., YAMAUCHI M., SEKI N., SUGAWARA T., SAITO T., MATSUDA Y.,  
RA ITOH H., NAGASE T., NOMURA N., HORI T.;  
RL GENOME RES. 6:439-447(1996).  
DR EMBL; X97186; E238806; -  
DR EMBL; D83243; G1304114; -  
FT CONFLICT 14 14 Y -> N (IN REF. 2).  
FT CONFLICT 295 295 L -> I (IN REF. 3).  
FT CONFLICT 471 471 N -> Y (IN REF. 3).

FT CONFLICT 967 967 E -> Q (IN REF. 3).  
FT CONFLICT 973 973 V -> L (IN REF. 3).  
FT CONFLICT 987 987 A -> V (IN REF. 3).  
SQ SEQUENCE 1427 AA; 154273 MW; 2DDEBF4 CRC32;

Query Match 75.0%; Score 48; DB 4; Length 1427;  
Best Local Similarity 60.0%; Pred. No. 4.27e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 688 TPPECTPVEN 697  
I:||||:;  
Qy 1 APPEDNPVED 10

RESULT 5  
ID O50383 PRELIMINARY; PRT; 129 AA.

AC O50383;  
DT 01-JUN-1998 (TREMREL. 06, CREATED)  
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 13.0 KD PROTEIN.  
GN MT004.11.  
OS MYCOBACTERIUM TUBERCULOSIS.  
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA OLIVER K., SKELTON J., BADCOCK K., CHURCHER C.M., HARRIS D.;  
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RX MEDLINE; 96181548.  
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELIA L.,  
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
RA COLE S.T.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).  
DR EMBL; AL009198; E1202271; -  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 129 AA; 12988 MW; 17B14A83 CRC32;

Query Match 73.4%; Score 47; DB 2; Length 129;  
Best Local Similarity 60.0%; Pred. No. 6.99e+00;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 28 APAQANPVD 37  
I:||||:;  
Qy 1 APPEDNPVED 10

RESULT 6  
ID O15392 PRELIMINARY; PRT; 142 AA.

AC O15392;  
DT 01-JAN-1998 (TREMREL. 05, CREATED)  
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
DE APOPTOSIS INHIBITOR SURVIVIN.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94148797.  
RA ALTIERI D.C.;  
RL J. BIOL. CHEM. 269:3139-3142(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95034823.

```

Db      658 PEDNPVE 664
      |||||
Qy      3 PEDNPVE 9

RESULT      9
ID Q92728      PRELIMINARY;      PRT;      53 AA.
AC Q92728:
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE RETINOBLASTOMA SUSCEPTIBILITY PROTEIN (RB).
GN RBL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
  [1]
RN R1
RP SEQUENCE FROM N.A.
RX MEDLINE; 88320373.
RA LEE E.Y., BOOKSTEIN R., YOUNG L.J., LIN C.J., ROSENFELD M.G.,
RA LEE W.H.;
RA LEE W.H.;
RL PROC. NATL. ACAD. SCI. U.S.A. 85:6017-6021(1988).
DR EMBL; M19701; G190968; -
SQ SEQUENCE 53 AA; 5392 MW; 2D5C3E39 CRC32;

Query Match      70.3%; Score 45; DB 4; Length 53;
Best Local Similarity 50.0%; Pred. No. 1.83e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db      27 PPEDEPQD 36
      ||||:|:|
Qy      1 APPEDNPVED 10

RESULT      10
ID P74580      PRELIMINARY;      PRT;      86 AA.
AC P74580:
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 9.9 KD PROTEIN.
OS SYNECHOCYSTIS SP.
OC EUBACTERIA; CYANOBACTERIA; CHROCOCCOCALES; SYNECHOCYSTIS.
  [1]
RN R1
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RA TABATA S.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
  [2]
RN R1
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RL DNA RES. 3:109-136(1996).
DR EMBL; D90916; G1653776; -
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 86 AA; 9909 MW; B95C9013 CRC32;

Query Match      70.3%; Score 45; DB 2; Length 86;
Best Local Similarity 62.5%; Pred. No. 1.83e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db      49 PDDTPVEE 56
      |:| |||
Qy      3 PEDNPVED 10

RESULT      11
ID Q26168      PRELIMINARY;      PRT;      97 AA.
AC Q26168:

```



DT 01-NOV-1996 (TREMREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
 DE DNA FRAGMENT WITH HIGHLY REPEATED SEQUENCE (FRAGMENT)  
 OS PLASMODIUM VIVAX.  
 OC EUKARYOTA; PROTOZOA; APICOMPLEXA; SPOOROZOA; COCCIDIA; EUCCOCCIDIIDA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA CAMPBELL J.R., FRANK E.D.;  
 RL SUBMITTED (APR-1989) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; X15129; G10087; -  
 FT NON-TER 1  
 SQ SEQUENCE 97 AA; 9736 MW; 3866D3C5 CRC32;

Query Match 70.3%; Score 45; DB 5; Length 97;  
 Best Local Similarity 71.4%; Pred. No. 1.83e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 13 PPEENPI 19  
 |||:|:|:  
 QY 2 PPEDNPV 8

RESULT 12  
 ID O33339 PRELIMINARY; PRT; 258 AA.  
 AC O33339;  
 DT 01-JAN-1998 (TREMREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)  
 DE DEHYDROGENASE.  
 GN MTV003.03C.  
 OS MYCOBACTERIUM TUBERCULOSIS.  
 OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA MURPHY L., HARRIS D.;  
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,  
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
 RA COLE S.T.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).  
 DR EMBL; AL008883; E1172949; -  
 DR PFAM; PF00106; adh\_short.  
 DR PFAM; PF00678; adh\_short.C2.  
 SQ SEQUENCE 258 AA; 26804 MW; 60151FCC CRC32;

Query Match 70.3%; Score 45; DB 2; Length 258;  
 Best Local Similarity 60.0%; Pred. No. 1.83e+01;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 93 SPEDNLIEN 102  
 :||||:|:  
 QY 1 APPEDNPVED 10

RESULT 13  
 ID O16657 PRELIMINARY; PRT; 287 AA.  
 AC O16657;  
 DT 01-JAN-1998 (TREMREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMREL. 05, LAST ANNOTATION UPDATE)  
 DE C03H5.3 PROTEIN.  
 GN C03H5.3.

OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE; 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
 RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
 RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
 RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,  
 RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
 RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
 RA RIFKIN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,  
 RA SONNHAMMER E., STADEN R., SULSTON J., THERRY-MIEG J., THOMAS K.,  
 RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
 RA WILKINSON-SPROAT J., WOHLDMANN P.;  
 RL NATURE 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA ROHLFING T., WOHLDMANN P.;  
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA WATERSTON R.;  
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; AF016674; G2315689; -  
 SQ SEQUENCE 287 AA; 31998 MW; 4D4ED5ED CRC32;

Query Match 70.3%; Score 45; DB 5; Length 287;  
 Best Local Similarity 50.0%; Pred. No. 1.83e+01;  
 Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 80 APSOENSVD 89  
 |||:|:|:|:  
 QY 1 APPEDNPVED 10

RESULT 14  
 ID O15319 PRELIMINARY; PRT; 317 AA.  
 AC O15319;  
 DT 01-JAN-1998 (TREMREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)  
 DE RNA POLYMERASE III SUBUNIT.  
 GN RPC39.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97315201.  
 RA WANG Z., ROEDER R.G.;  
 RL GENES DEV. 11:1315-1326(1997).  
 DR EMBL; U93869; G2228752; -  
 SQ SEQUENCE 317 AA; 36144 MW; BCC34DCF CRC32;

Query Match 70.3%; Score 45; DB 4; Length 317;  
 Best Local Similarity 62.5%; Pred. No. 1.83e+01;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 10 PPADNPVE 17  
 ||:|:|:  
 QY 2 PPEDNPVE 9

RESULT 15  
 ID Q56796 PRELIMINARY; PRT; 339 AA.  
 AC Q56796;  
 DT 01-NOV-1996 (TREMREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMREL. 01, LAST ANNOTATION UPDATE)

DE JUGLANDIS COPPER-RESISTANCE GENES, COMPLETE CDS.  
OS XANTHOMONAS CAMPESTRIS.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
OC PSEUDOMONADACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-COPPER-RESISTANCE;  
RX MEDLINE; 94110224.  
RA LEE Y.A., HENDSON M., PANOPOULOS N.J., SCHROTH M.N.;  
RL J. BACTERIOL. 176:173-188(1994).  
DR EMBL; L19222; G461142; -.  
SQ SEQUENCE 339 AA; 37041 MW; D1400210 CRC32;

Query Match 70.3%; Score 45; DB 2; Length 339;  
Best Local Similarity 60.0%; Pred. NO. 1.83e+01;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 121 PPSDHPVHD 130  
QY 1 APPEDNEVED 10

Search completed: Thu May 13 15:29:42 1999  
Job time : 27 secs.

\*\*\*\*\*

W A S R L A

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MParch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 13 15:31:38 1999; Maspar time 2.52 Seconds

Tabular output not generated. 64.113 Million cell updates/sec

Title: >US-09-040-485-7  
Description: (1-10) from US09040485.pep  
Perfect Score: 66  
Sequence: 1 BEQEVPPDT 10

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq32  
1:part1 2:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 14.792; Variance 49.512; scale 0.299

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description                  | Pred. No. |
|------------|-------|-------------|--------|----|------------------------------|-----------|
| 1          | 45    | 68.2        | 3144   | 29 | W44742 Human huntingtin prot | 1.72e+02  |
| 2          | 45    | 68.2        | 3144   | 11 | R58777 Protein encoded by Hu | 1.72e+02  |
| 3          | 45    | 68.2        | 3144   | 26 | W36887 Previously undescribe | 1.72e+02  |
| 4          | 45    | 68.2        | 3144   | 22 | W09871 Human huntingtin      | 1.72e+02  |
| 5          | 43    | 65.2        | 140    | 1  | R05669 Gamma-chicken atrial  | 2.72e+02  |
| 6          | 43    | 65.2        | 414    | 16 | R30355 Transforming growth f | 2.72e+02  |
| 7          | 43    | 65.2        | 414    | 14 | R73597 Human TGF-beta 2 prot | 2.72e+02  |
| 8          | 43    | 65.2        | 442    | 1  | P91899 Sequence encoded by h | 2.72e+02  |
| 9          | 43    | 65.2        | 442    | 16 | R79922 Human transforming gr | 2.72e+02  |
| 10         | 43    | 65.2        | 442    | 4  | R20125 Sequence of human tra | 2.72e+02  |
| 11         | 43    | 65.2        | 442    | 1  | R05718 Human TGF-Beta2-442 p | 2.72e+02  |
| 12         | 43    | 65.2        | 700    | 1  | R03663 Human myb related gen | 2.72e+02  |
| 13         | 42    | 63.6        | 193    | 25 | W23619 Prolactin antagonist  | 3.40e+02  |
| 14         | 42    | 63.6        | 225    | 1  | R82078 Recombinant rat prepr | 3.40e+02  |
| 15         | 42    | 63.6        | 226    | 3  | R14599 Rat prolactin         | 3.40e+02  |
| 16         | 42    | 63.6        | 392    | 2  | R12345 Toxoplasma gondii pro | 3.40e+02  |
| 17         | 42    | 63.6        | 428    | 2  | R12352 Toxoplasma gondii P66 | 3.40e+02  |
| 18         | 42    | 63.6        | 473    | 22 | W14005 Human SHC protein.    | 3.40e+02  |

|    |    |      |      |    |        |                       |          |
|----|----|------|------|----|--------|-----------------------|----------|
| 19 | 42 | 63.6 | 473  | 15 | R84637 | Shc protein.          | 3.40e+02 |
| 20 | 42 | 63.6 | 474  | 17 | R97243 | SHC phosphotyrosine b | 3.40e+02 |
| 21 | 42 | 63.6 | 494  | 15 | R76615 | Saccharomyces cerevis | 3.40e+02 |
| 22 | 42 | 63.6 | 494  | 6  | R31888 | Defective tyrosine ki | 3.40e+02 |
| 23 | 42 | 63.6 | 494  | 11 | R58519 | Protein kinase (HRR25 | 3.40e+02 |
| 24 | 41 | 62.1 | 570  | 26 | W36002 | Human Fchd531 gene pr | 4.26e+02 |
| 25 | 41 | 62.1 | 1719 | 16 | R92100 | Human RIZ allele D283 | 4.26e+02 |
| 26 | 41 | 62.1 | 3119 | 29 | W44743 | Mouse huntingtin prot | 4.26e+02 |
| 27 | 41 | 62.1 | 3119 | 26 | W36888 | Mouse huntingtin's di | 4.26e+02 |
| 28 | 40 | 60.6 | 101  | 18 | R90022 | Ubi7 ubiquitin-lytic  | 5.32e+02 |
| 29 | 40 | 60.6 | 240  | 4  | R22597 | Poetal oncogene Pem s | 5.32e+02 |
| 30 | 40 | 60.6 | 463  | 25 | W19394 | Human calpastatin (cl | 5.32e+02 |
| 31 | 40 | 60.6 | 673  | 1  | R05701 | Human calpastatin pol | 5.32e+02 |
| 32 | 40 | 60.6 | 706  | 18 | R98925 | Human APLP2.          | 5.32e+02 |
| 33 | 40 | 60.6 | 706  | 9  | R47499 | Human APLP2.          | 5.32e+02 |
| 34 | 40 | 60.6 | 763  | 10 | R53778 | Sequence of human amy | 5.32e+02 |
| 35 | 40 | 60.6 | 1350 | 17 | R91296 | Drosophila nitric oxi | 5.32e+02 |
| 36 | 40 | 60.6 | 1618 | 5  | R27205 | Human nestin.         | 5.32e+02 |
| 37 | 40 | 60.6 | 1618 | 11 | R60127 | Human nestin protein  | 5.32e+02 |
| 38 | 40 | 60.6 | 2205 | 15 | R79048 | Infectious rubella vi | 5.32e+02 |
| 39 | 40 | 60.6 | 5035 | 5  | R25450 | MH mutant porcine rya | 5.32e+02 |
| 40 | 39 | 59.1 | 740  | 5  | R27530 | Plasmodium falciparum | 6.62e+02 |
| 41 | 39 | 59.1 | 858  | 1  | P80806 | Sequence of env prote | 6.62e+02 |
| 42 | 39 | 59.1 | 891  | 1  | P82677 | ENVRN sequence from H | 6.62e+02 |
| 43 | 39 | 59.1 | 957  | 11 | R56990 | Bacillus deramificans | 6.62e+02 |
| 44 | 39 | 59.1 | 1717 | 27 | W23331 | Neuroblastoma indicat | 6.62e+02 |
| 45 | 39 | 59.1 | 1831 | 27 | W23329 | Microtubule-associate | 6.62e+02 |

## ALIGNMENTS

RESULT 1  
ID W44742 standard; Protein; 3144 AA.  
AC W44742;

DT 01-JUN-1998 (first entry)

DE Human huntingtin protein.

KW Human; huntingtin gene; Huntington's disease; chromosome; marker;  
KW locus; antisense; gene therapy; diagnosis.

OS Homo sapiens.

PN US5693757-A.

PD 02-DEC-1997.

PF 30-MAY-1995; 453265.

PR 20-MAY-1994; US-246982.

PR 05-MAR-1993; US-027498.

PR 01-JUL-1993; US-085000.

PR 30-MAY-1995; US-453265.

PA (GHEO ) GEN HOSPITAL CORP.

PI Ambrose CM, Duyao MP, Gusella JF, MacDonald ME;

DR WPI; 98-031815/03.

DR N-PSDB; V05828.

PT Huntingtin protein and related nucleic acid - for diagnosis or

PT therapy of Huntington's disease

PS Claim 2; Fig 4; 112pp; English.

CC This is the amino acid sequence of the human huntingtin protein.

CC The gene sequence is characterised in that it contains a number of

CC CAG repeats in the 5' region (in this case 23 repeats). In healthy

CC individuals the usual number of CAG repeats in the gene is 11-34,

CC whereas in patients suffering from Huntington's disease (HD), the

CC number of repeats increases to 37-73 or 37-86. The huntingtin gene

CC spans 210 kb and encodes a protein of 348 kD. The gene is found in a

CC 500 kb region between the chromosomal markers D4S180 and D4S182 and is

CC preferentially mapped to the locus 4p16.3. The protein or the gene

CC encoding it, is useful for detecting a predisposition to develop HD,

CC for diagnosis and treatment of HD, especially by antisense and gene

CC therapy.

CC Sequence 3144 AA;

Query Match 68.2%; Score 45; DB 29; Length 3144;

Best Local Similarity 60.0%; Pred. No. 1.72e+02;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 eeeevdpnt 2352

|||||

```

Oy 1 EEQEQVPPDT 10
RESULT 2
ID R58777 standard; Protein; 3144 AA.
AC R58777;
DE 13-APR-1995 (first entry)
DE Protein encoded by Huntingtin DNA/IT15 gene.
KW Polymerase chain reaction; primer; PCR; amplif; Huntington; IT15;
KW open reading frame; polymorphic; (CAG)n; trinucleotide repeat;
KW allele; Huntington's disease; HD; chromosome; juvenile HD; unstable;
KW expandable; linkage disequilibrium.
OS Homo sapiens.
PN EP-614977-A.
PD 14-SEP-1994. 301587.
PF 07-MAR-1994. 301587.
PR 05-MAR-1993; US-027498.
PR 01-JUL-1993; US-085000.
PA (GEHO ) GEN HOSPITAL CORP.
PI Ambrose CM, Duyao MP, Gusella JF, MacDonald ME;
DR N-PSDB; Q67401.
DR WPI; 94-281205/35.
PT New huntingtin protein and related nucleic acid, antibodies etc.
PT - for treatment and diagnosis of neuro-degenerative disease,
PT specifically Huntington's disease
PS Claim 1; Page 22-33; 66pp; English.
CC This sequence is encoded by a large gene, termed "Huntingtin" or
CC "IT15" which is present in the proximal part of the 0.5 mb segment
CC between members D4S180 and D4S182 on chromosome 4. The IT15 gene
CC spans about 210 kb and encodes a protein of approx. 348 kb. The
CC huntingtin open reading frame contains a polymorphic (CAG)n
CC trinucleotide repeat with at least 17 alleles in the normal
CC population, varying from about 11 to about 34 CAG copies. On
CC Huntington's disease (HD) chromosome, the length of the CAG repeat
CC is substantially increased, with between 37 to at least 73 copies.
CC This shows an apparent correlation with age of onset, the longest
CC segments are detected in juvenile HD cases. The presence of an
CC unstable, expandable trinucleotide repeat on HD chromosomes in the
CC region of the strongest linkage disequilibrium with the disorder
CC suggest that this alteration underlies the dominant phenotype of HD.
SQ Sequence 3144 AA;

Query Match 68.2%; Score 45; DB 11; Length 3144;
Best Local Similarity 60.0%; Pred. No. 1.72e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 eeeevdpnt 2352
Oy 1 EEQEQVPPDT 10

RESULT 3
ID W36887 standard; Protein; 3144 AA.
AC W36887;
DE 13-MAR-1998 (first entry)
DE Previously undescribed protein encoded by a novel huntingtin (IT15) gene.
KW Huntingtin gene; IT15 gene; Huntington's disease; trinucleotide repeat;
KW neurodegenerative disorder; HD; gene therapy.
OS Homo sapiens.
PN US568628-A.
PD 11-NOV-1997. 246982.
PF 20-MAY-1994. 246982.
PR 05-MAR-1993; US-027498.
PR 01-JUL-1993; US-085000.
PA (GEHO ) GEN HOSPITAL CORP.
PI Ambrose CM, Duyao MP, Gusella JF, MacDonald ME;
DR N-PSDB; Q79924.
DR Nucleic acid encoding huntingtin protein - useful for gene therapy
PT of Huntington's disease
PS Claim 2; Fig 4; 112pp; English.
CC The present sequence represents a previously undescribed protein,
CC encoded by a novel gene, termed huntingtin or IT15. The huntingtin

```

---

```

CC reading frame contains a polymorphic (CAG)n trinucleotide repeat with at
CC least 17 alleles in the normal population, varying from about 11 to 34
CC CAG copies. Huntington's disease (HD) is a progressive neurodegenerative
CC disorder characterised by motor disturbance, cognitive loss and
CC psychiatric manifestations. The genetic defect causing HD is assigned to
CC chromosome 4. On HD chromosomes, the length of the trinucleotide CAG
CC repeat is substantially increased, e.g. about 37 to at least 73 copies.
CC The huntingtin gene and proteins encoded by it, may be used for the
CC diagnosis or treatment of Huntington's disease. The huntingtin gene
CC is especially used in gene therapy of a symptomatic or presymptomatic
CC patient. The method comprises providing a functional huntingtin gene with
CC a (CAG)n repeat of the normal range of 11-34 copies, or an antisense
CC sequence, to the desired cells of the patient, in a manner that permits
CC the expression of the huntingtin protein provided by the gene, or
CC inhibits expression of the mutated huntingtin gene, for a time and in a
CC quantity sufficient to provide the huntingtin function to the cells of
CC the patient.
SQ Sequence 3144 AA;

Query Match 68.2%; Score 45; DB 26; Length 3144;
Best Local Similarity 60.0%; Pred. No. 1.72e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 eeeevdpnt 2352
Oy 1 EEQEQVPPDT 10

RESULT 4
ID W09871 standard; Protein; 3144 AA.
AC W09871; 1997 (first entry)
DE Human huntingtin.
KW Huntingtin associated protein-1; HAP1; Huntington's disease.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Protein 1..230
FT W09717443-A1.
FT W09717443-A1.
PN 15-MAY-1997.
PF 08-NOV-1996; U17858.
PR 09-NOV-1995; US-556419.
PA (UUCO ) UNIV JOHNS HOPKINS.
PI Lanahan A, Li S, Li X, Ross CA, Sharp AH, Snyder S;
PI Worley PF;
DR WPI; 97-281032/25.
PT Determination of the binding of huntingtin to huntingtin-associated
PT protein-1 - useful for screening for drugs for treating or
PT preventing Huntington's disease
PS Claim 20; Page 22-32; 69pp; English.
CC Human huntingtin (Hn) polypeptide (W09871) is the product of the
CC Huntington's disease (HD) locus. Proteins which specifically bind
CC to Hn, such as human huntingtin associated protein-1 (HAP1) (see
CC also W09870), can be used in assays for screening drug candidates.
CC The binding between Hn and HAP1 is enhanced by an expanded
CC polyglutamine repeat in Hn, the length of which correlates with the
CC time of disease onset. HAP1, in contrast to Hn, is expressed
CC selectively in the brain, suggesting that it may contribute to the
CC brain-specific pathology of HD. Hn, or portions of it, esp. amino
CC acids 1-230, or yeast cells expressing Hn, can be used to identify
CC cpds. that bind to, displace or prevent binding of Hn and HAP1.
CC The method is useful for screening candidate drugs for treating,
CC delaying onset of, or preventing HD.
SQ Sequence 3144 AA;

Query Match 68.2%; Score 45; DB 22; Length 3144;
Best Local Similarity 60.0%; Pred. No. 1.72e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 eeeevdpnt 2352
Oy 1 EEQEQVPPDT 10

```

```

RESULT 5
ID R05669 standard; peptide; 140 AA.
AC R05669;
DT 15-AUG-1990 (first entry)
DE Gamma-chicken atrial natriuretic peptide.
KW Gamma-chicken atrial natriuretic peptide; diuretic; hypertensive.
OS Gallus sp.
FH Key Location/Qualifiers
FT disulfide_bond 118..134
FT region 25..140
FT /label=Gamma-chANP specific.
PN J02025499-A.
PD 26-JAN-1990.
PF 14-JUL-1988; 173739.
PR 14-JUL-1988; JP-173739.
PA (MATSU) Matsuo T.
DR WPI: 90-071804/10.
DR N-PSDB; Q03465.
PT Physio-active peptide derived from birds - has cysteine bridge,
PT specified amino acid sequence and diuretic and hypertensive properties.
PS Disclosure; Fig 13; 16pp; Japanese.
CC Gamma-chANP is obtained from the 12 kD fraction obtained from treated
CC homogenised chicken heart tissue.
CC See also Q03466-Q03468, R03301 and R03302.
SQ Sequence 140 AA;

Query Match 65.2%; Score 43; DB 1; Length 140;
Best Local Similarity 44.4%; Pred. No. 2.72e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 66 qtgeippe 74
QY 1 EEOQEVPPD 9

RESULT 6
ID R83055 standard; protein; 414 AA.
AC R83055;
DT 25-JUN-1996 (first entry)
DE Transforming growth factor-beta 2.
KW macrophage inducible nitric oxide synthase; iNOS; constitutive NOS;
KW interleukin-1-beta; transforming growth factor-beta; TGF-beta; IL1-beta;
KW nitric oxide production; hypotension; inflammation; septic shock;
KW treatment.
OS Mammalian sp.
FH Key Location/Qualifiers
FT protein 303..414
FT /note= "represents the mature active TGF beta-1 mol."
PN W09526745-A1.
PD 12-OCT-1995.
PF 05-APR-1994; U03705.
PR 05-APR-1994; WO-U03705.
PA (HARD ) HARVARD COLLEGE.
PI Lee M, Perrella MA;
DR WPI: 95-35843/46.
DR N-PSDB; T05877.
PT Treatment of hypotension, esp. in septic shock - by administering
PT transforming growth factor-beta e.g. to inhibit inducible nitric
PT oxide synthase gene transcription
PS Disclosure; Fig 18; 52pp; English.
CC Transforming growth factor-beta 2 (TGF-beta 2) has been found to inhibit
CC inducible nitric oxide synthase (iNOS) gene transcription, esp. in
CC interleukin-1-beta (IL1-beta) stimulated rat smooth muscle cells, and at
CC a dose which does not inhibit constitutive NOS. TGF-beta 1 (R83054) or 2
CC or their active fragments (esp. derived from the carboxy-terminal 112
CC amino acids), can be used in the treatment of hypotension, such as that
CC associated with severe inflammation or septic shock.
SQ Sequence 414 AA;

Query Match 65.2%; Score 43; DB 16; Length 414;
Best Local Similarity 62.5%; Pred. No. 2.72e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 59 epeevppe 66
QY 2 EEOQEVPPD 9

RESULT 7
ID R73597 standard; protein; 414 AA.
AC R73597;
DT 20-DEC-1995 (first entry)
DE Human TGF-beta 2 protein.
KW Transforming growth factor-beta; Human TGF-beta protein; TGF-beta 1;
KW TGF-beta 2; TGF-beta 3; osteogenic cell source; OCS; bone deficiency;
KW bone-inducing cofactor.
OS Homo sapiens.
PN US5409896-A.
PD 25-APR-1995.
PF 12-NOV-1993; 401906.
PR 01-SEP-1989; US-401906.
PR 12-NOV-1991; US-790856.
PR 18-MAY-1993; US-063841.
PR 12-NOV-1993; US-132405.
PA (GETH ) GENENTECH INC.
PI Ammann AJ, Rudman CG;
DR WPI: 95-169610/22.
PT Compsn. for treating skeletal tissue deficiency - comprising
PT transforming growth factor-beta and an osteogenic cell source in a
PT carrier
PS Disclosure; Column 17-20; 19pp; English.
CC This sequence represents human transforming growth factor-beta 2
CC (TGF-beta 2). The sequences for human TGF-beta 1 (see R73596) and human
CC TGF-beta 3 (see R73598) are claimed within the scope of the invention.
CC The invention is a composition consisting of a TGF-beta protein and an
CC osteogenic cell source (OCS) formulated in an acceptable carrier other
CC than a bone morphogenic cofactor. This composition can be used for the
CC restoration of bone deficiency. This provides for the generation of a
CC mature bone only where it is required, without the inclusion of a
CC specific bone-inducing cofactor. This method can be used with any of the
CC 5 human TGF-beta s or with TGF-beta from other species.
SQ Sequence 414 AA;

Query Match 65.2%; Score 43; DB 14; Length 414;
Best Local Similarity 62.5%; Pred. No. 2.72e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 59 epeevppe 66.
QY 2 EEOQEVPPD 9

RESULT 8
ID P91899 standard; protein; 442 AA.
AC P91899;
DT 23-DEC-1990 (first entry)
DE Sequence encoded by human transforming growth factor (TGF) beta-2
DE precursor 442 cDNA in pPC-21
KW Cell differentiation; cell proliferation.
OS Homo sapiens.
FH Key Location/Qualifiers
FT region 20...442
FT /note="Claimed"
FT peptide 4..19
FT /note="Signal"
FT cleavage_site 20..21
FT region 116..144
FT /note="This entire SQ is replaced with Asn in simian
FT TGF-beta-2-414"
FT modified_site 72
FT /note="Potential glycosylation site"
FT modified_site 168
FT /note="Potential glycosylation site"
FT modified_site 269
FT /note="Potential glycosylation site"

```

FT protein 331..442  
 PN DE3833897-A.  
 PD 03-MAY-1989.  
 PF 05-OCT-1988; 833897.  
 PR 18-AUG-1988; US-234065.  
 PA (ONCO-) Oncogen.  
 PI Purchio AF, Madisen L, Webb N;  
 DR WPI; 89-138796/19.  
 DR N-PSDB; N90767.  
 PT New DNA sequence encoding transforming growth factor beta 2 -  
 used for large scale expression in eucaryotic cells  
 PS Claim 4; Fig 1a; 2pp; German.  
 CC PolyA-RNA was isolated from the tamoxifen-treated, human prostatic  
 adenocarcinoma line PC-3 and converted to cDNA. TGF DNA is pref. used  
 for control of the SV40 promoter. and expressed in CHO cells. The simian  
 SO is also claimed.  
 SQ Sequence 442 AA;

Query Match 65.2%; Score 43; DB 1; Length 442;  
 Best Local Similarity 62.5%; Pred. No. 2.72e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 59 epeevppe 66  
 | :||||:  
 Qy 2 EQQEVPPD 9

RESULT 9  
 ID R79922 standard; Protein; 442 AA.  
 AC R79922;  
 DT 28-MAY-1996 (first entry)  
 DE Human transforming growth factor-2.  
 KW TGF-beta1; TGF-beta2; transforming growth factor; protein;  
 KW cell differentiation; cell proliferation; CHO; Chinese hamster;  
 KW ovary; COS; monkey kidney; animal; mammal.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT peptide 4..19  
 FT cleavage\_site 20..21  
 FT /note= "signal peptide"  
 FT peptide 331..442  
 FT /note= "putative signal sequence cleavage site"  
 FT peptide 331..442  
 FT /note= "mature peptide"  
 PN EP-676474-A1.  
 PD 11-OCT-1995.  
 PF 14-DEC-1989; 104223.  
 PR 16-DEC-1988; US-285140.  
 PR 05-DEC-1989; US-446020.  
 PA (ONCO ) ONCOGEN LP.  
 PI Madisen L, Purchio AF, Webb N;  
 DR WPI; 95-346094/45.  
 DR N-PSDB; T04116.  
 PT Hybrid transforming growth factor beta-1/TGF-beta-2 precursor - used  
 to produce biologically active, mature TGF-beta-2  
 PS Disclosure; Fig.1a; 52pp; English.  
 CC This sequence is expressed in a host cell, preferably a  
 COS or CHO cell, so the host cell produces active TGF-beta2. The  
 CC produced TGF-beta2 protein can be used to regulate cellular  
 CC differentiation and proliferation.  
 SQ Sequence 442 AA;

Query Match 65.2%; Score 43; DB 16; Length 442;  
 Best Local Similarity 62.5%; Pred. No. 2.72e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 59 epeevppe 66  
 | :||||:  
 Qy 2 EQQEVPPD 9

RESULT 10  
 ID R20125 standard; Protein; 442 AA.  
 AC R20125;

DT 16-APR-1992 (first entry)  
 DE Sequence of human transforming growth factor (TGF) beta-2-442.  
 KW Hypertension therapy; hypotensive agent; blood pressure modulator..  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT peptide 4..19  
 FT protein 330..442  
 PN WO9119513-A.  
 PD 26-DEC-1991.  
 PF 20-JUN-1991; U04449.  
 PR 20-JUN-1990; US-541221.  
 PA (BRIM ) BRISTOL-MYERS SQUIB.  
 PI Oleson FB, Comerreski CR;  
 DR WPI; 92-024199/03.  
 DR N-PSDB; Q20290.  
 PT Use of transforming growth factor (TGF)-beta and their  
 PT antagonists - for modulating blood pressure, for treating  
 PT hypertension and hypotension  
 PS Disclosure; Fig 2; 42pp; English.  
 CC A new method for treating hypertension comprises administering a  
 CC transforming growth factor (TGF)-beta to an individual at a dose  
 CC effective for lowering blood pressure; the TGF-beta may be e.g.  
 CC mature TGF-beta, TGF-beta2, a mature TGF-beta2 hybrid, TGF-  
 CC beta1 precursor, a latent TGF-beta2 precursor, hybrid TGF-beta1/TGF-  
 CC beta2 precursor, a latent TGF-beta1 complex or a latent TGF-beta2  
 CC complex.  
 SQ Sequence 442 AA;

Query Match 65.2%; Score 43; DB 4; Length 442;  
 Best Local Similarity 62.5%; Pred. No. 2.72e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 59 epeevppe 66  
 | :||||:  
 Qy 2 EQQEVPPD 9

RESULT 11  
 ID R05748 standard; protein; 442 AA.  
 AC R05748;  
 DT 02-NOV-1990 (first entry)  
 DE Human TGF-Beta2-442 precursor.  
 KW Human TGF-Beta2 precursor; cancer; tumorcide; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT protein 331..442  
 FT peptide 4..19  
 FT /label-Sinal peptide.  
 PN EP-376785-A.  
 PD 4-JUL-1990.  
 PF 14-DEC-1989; 403480.  
 PR 16-DEC-1988; US-285140.  
 PR 5-DEC-1989; US-446020.  
 PA (ONCO ) Oncogen Ltd Partner.  
 PI Purchio AF, Madisen L, Webb N;  
 DR WPI; 90-203127/27.  
 DR N-PSDB; Q05126  
 PT Cloning and expression of transforming growth factor beta 2 -  
 used for treatment of tumors or for augmenting wound healing.  
 PS Claim 1; Fig 1a; 58pp; English.  
 CC TGF-Beta2 may be used in treatment of tumors at effective doses,  
 CC and may also be useful in augmenting wound healing by stimulating  
 CC cell proliferation. The growth factor can be produced at high  
 CC levels from a CHO expression system.  
 SQ Sequence 442 AA;

Query Match 65.2%; Score 43; DB 1; Length 442;  
 Best Local Similarity 62.5%; Pred. No. 2.72e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 59 epeevppe 66  
 | :||||:  
 Qy 2 EQQEVPPD 9

RESULT 12  
 ID R03663 standard; protein; 700 AA.  
 AC R03663;  
 DT 21-AUG-1990 (first entry)  
 DE Human myb related gene product.  
 KW Cancer; myb; myeloblast; ds.  
 OS Homo sapiens.  
 PN J02053486-A.  
 PD 22-FEB-1990.  
 PF 19-AUG-1988; 205907.  
 PR 19-AUG-1988; JP-205907.  
 FA (RIKA) Rikagaku Kenkyusho.  
 DR WPI; 90-103118/14.  
 DR N-PSDB; Q03741.  
 PT Human Myeloblast related gene -  
 coded by bases sequence including 105 adenine 1150 thymine, etc.  
 PS Claim 2; Fig 2; 8pp; Japanese.  
 CC Proteins generated by the sequence may be used to raise antibodies useful  
 in determining the copy number of the cancer gene.  
 SQ Sequence 700 AA;

Query Match 65.2%; Score 43; DB 1; Length 700;  
 Best Local Similarity 40.0%; Pred. No. 2.72e+02;  
 Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 277 d0qegsppet 286  
 : : : : :  
 QY 1 EEQVEVPPDT 10

RESULT 13  
 ID W23619 standard; Protein; 193 AA.  
 AC W23619;  
 DT 11-FEB-1998 (first entry)  
 DE Prolactin antagonist (substituted rat prolactin).  
 KW Prolactin antagonist; phosphorylation; hyperprolactinaemia;  
 prolactinoma; prostate cancer; tumour; T-lymphoma; infertility;  
 lactation; miscarriage; ovulation; antibody; therapy; rat.  
 OS Rattus sp.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Misc\_difference 173 /note= "variable site"  
 FT  
 PN W09727865-A1.  
 PD 07-AUG-1997.  
 PF 30-JAN-1997; U01435.  
 PR 31-JAN-1996; US-594809.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Walker AM.  
 DR WPI; 97-402308/37.  
 PT Substituted prolactin peptide(s) and proteins having an amino acid  
 substitution for serine in the C-terminal - useful as prolactin  
 antagonists, e.g. for treating prolactin dependent cancers  
 PS Claim 4; Page 101-102; 158pp; English.  
 CC This protein comprises rat prolactin, substituted at residue 173  
 (serine in the native sequence). It has prolactin antagonist  
 activity, antagonising the stimulation of T lymphoma cell growth in  
 the presence of non-phosphorylated prolactin. Other claimed  
 CC prolactin antagonists (see W23608-18) comprise prolactin  
 CC substitution mutant C-terminal peptides. Claimed antagonists can  
 be used for the treatment of prolactin dependent cancers and can  
 inhibit T-lymphoma cell proliferation. They are also useful for  
 CC treatment of prolactinoma, infertility related to abnormal  
 CC prolactin regulation, some forms of prostatic cancer, miscarriage  
 CC and ovulation irregularities, as well as in assays to measure  
 CC levels of non-phosphorylated and phosphorylated prolactin as an  
 CC indicator of reproductive pathologies and presence or status of a  
 CC prolactin-dependent tumour, and to raise polyclonal and monoclonal  
 CC antibodies.  
 SQ Sequence 193 AA;

Query Match 63.6%; Score 42; DB 25; Length 193;  
 Best Local Similarity 55.6%; Pred. No. 3.40e+02;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 64 eqagkvpppe 72  
 : : : : :  
 QY 1 EEQVEVPPD 9

RESULT 14  
 ID P82078 standard; protein; 225 AA.  
 AC P82078;  
 DT 18-OCT-1990 (first entry)  
 DE Recombinant rat preprolactin.  
 KW Prolactin; milk; contraceptive; dairy cows; lactation.  
 OS Rattus rattus.  
 FH Key Location/Qualifiers  
 FT protein 1..225  
 FT /label=preprolactin 29..225  
 FT /label=prolactin  
 PN US4725549-A.  
 PD 16-FEB-1988.  
 PF 23-MAR-1984; 592714.  
 PR 22-SEP-1980; US-189160.  
 PR 23-MAR-1984; US-592714.  
 PA (REGC) University of California.  
 PI Cooke NE; Baxter JD;  
 DR WPI; 88-070922/10.  
 DR N-PSDB; N80114.  
 PT DNA coding for prolactin - obtd. by prepn. of reverse transcript  
 of mRNA coding for prolactin and inserting into a transfer vector.  
 PS Disclosure; p; English.  
 CC The cDNA encoding the prolactin can be inserted into expression vectors  
 for the prodn. of rat prolactin which can be admin. to dairy cows to  
 CC increase milk yield. The protein can also be used as a female  
 CC contraceptive and to ensure adequate milk prodn. for breast feeding  
 CC mothers.  
 CC See also P82079.  
 SQ Sequence 225 AA;

Query Match 63.6%; Score 42; DB 1; Length 225;  
 Best Local Similarity 55.6%; Pred. No. 3.40e+02;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 96 eqagkvpppe 104  
 : : : : :  
 QY 1 EEQVEVPPD 9

RESULT 15  
 ID R14599 standard; Protein; 226 AA.  
 AC R14599;  
 DT 21-JAN-1992 (first entry)  
 DE Rat prolactin.  
 KW Recombinant DNA.  
 OS Rattus rattus.  
 PN J03219876-A.  
 PD 27-SEP-1991.  
 PF 24-JAN-1990; 014511.  
 PR 24-JAN-1990; JP-014511.  
 PA (SHIK-) SHIKISHIMA BOSEKI K.  
 DR WPI; 91-329111/45.  
 DR N-PSDB; Q14451, Q14452.  
 PT Mass-prodn. of rat prolactin-producing recombinant DNA - is by  
 integration of promoter, Shine-Dalgarno sequence and translation  
 PT initiation codon upstream of table 1 coding gene.  
 PS Disclosure; Table 1; 12pp; Japanese.  
 CC The gene encoding the protein can be ligated into an expression  
 CC plasmid with a promoter, SD sequence and initiation codon for the  
 CC prodn. of recombinant rat prolactin. The protein is useful for the  
 CC study of the physiological activity of rat and human prolactin.  
 SQ Sequence 226 AA;

Query Match 63.6%; Score 42; DB 3; Length 226;  
Best Local Similarity 55.6%; Pred. No. 3.40e+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 97 eqagkvppe 105  
! : ! ! ! !  
QY 1 EEQOEVPDP 9

Search completed: Thu May 13 15:31:55 1999  
Job time : 17 secs.



\*\*\*\*\*

W P S R L H  
(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 13 15:32:14 1999; MasPar time 2.87 Seconds

Tabular output not generated. 130.735 Million cell updates/sec

Title: >US-09-040-485-7

Description: (1-10) from US09040485.pap

Perfect Score: 66  
Sequence: 1 EEQQEVPPTD 10

Scoring table: PAM 150

Gap 15

Searched: 116738 seqs, 37463448 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: pir58

1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 21.325; Variance 28.553; scale 0.747

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match % | ID   | Description | Pred. No. |
|------------|-------|---------------|------|-------------|-----------|
| 1          | 65    | 98.5          | 754  | 1 BABOH     | 1.90e-03  |
| 2          | 65    | 98.5          | 757  | 2 I38423    | 1.90e-03  |
| 3          | 55    | 83.3          | 816  | 2 S54518    | 2.96e-01  |
| 4          | 49    | 74.2          | 226  | 2 A49159    | 4.98e+00  |
| 5          | 47    | 71.2          | 147  | 2 H70305    | 1.22e+01  |
| 6          | 46    | 69.7          | 68   | 2 C22175    | 1.90e+01  |
| 7          | 46    | 69.7          | 167  | 2 S50808    | 1.90e+01  |
| 8          | 46    | 69.7          | 7962 | 2 I38346    | 1.90e+01  |
| 9          | 45    | 68.2          | 217  | 2 JC5728    | 2.93e+01  |
| 10         | 45    | 68.2          | 411  | 2 S47436    | 2.93e+01  |
| 11         | 45    | 68.2          | 852  | 2 A34373    | 2.93e+01  |
| 12         | 45    | 68.2          | 3144 | 2 A46068    | 2.93e+01  |
| 13         | 45    | 68.2          | 3924 | 2 S37431    | 2.93e+01  |
| 14         | 44    | 66.7          | 224  | 2 S37862    | 4.48e+01  |
| 15         | 44    | 66.7          | 354  | 2 H65206    | 4.48e+01  |
| 16         | 44    | 66.7          | 411  | 2 S65916    | 4.48e+01  |
| 17         | 44    | 66.7          | 414  | 2 JN0866    | 4.48e+01  |
| 18         | 44    | 66.7          | 700  | 1 S01991    | 4.48e+01  |
| 19         | 43    | 65.2          | 140  | 2 S14320    | 6.83e+01  |
| 20         | 43    | 65.2          | 184  | 2 S48035    | 6.83e+01  |
| 21         | 43    | 65.2          | 324  | 2 S58142    | 6.83e+01  |
| 22         | 43    | 65.2          | 370  | 2 I46054    | 6.83e+01  |
| 23         | 43    | 65.2          | 395  | 2 S38812    | 6.83e+01  |

24 43 65.2 405 2 A30357 retinal S-antigen - h 6.83e+01  
25 43 65.2 412 2 A39489 transforming growth f 6.83e+01  
26 43 65.2 414 1 WFMKB2 transforming growth f 6.83e+01  
27 43 65.2 414 2 A31249 transforming growth f 6.83e+01  
28 43 65.2 525 2 JN0059 hypothetical 57.4K pr 6.83e+01  
29 43 65.2 529 2 A56516 nuclear localization 6.83e+01  
30 43 65.2 529 2 S57873 pendulin - mouse 6.83e+01  
31 43 65.2 529 2 S57345 m-importin (nuclear p 6.83e+01  
32 43 65.2 572 1 HNN274 hemagglutinin-neurami 6.83e+01  
33 43 65.2 572 2 A46451 hemagglutinin-neurami 6.83e+01  
34 43 65.2 572 1 HNN282 hemagglutinin-neurami 6.83e+01  
35 43 65.2 572 1 HNN273 hemagglutinin-neurami 6.83e+01  
36 43 65.2 572 1 HNN273 hemagglutinin-neurami 6.83e+01  
37 43 65.2 855 2 A48168 proliferating-cell nu 6.83e+01  
38 43 65.2 971 2 S23408 prematurely terminate 6.83e+01  
39 43 65.2 1012 2 I53172 RAE-28 - mouse 6.83e+01  
40 43 65.2 1281 2 JC5368 dynactin 1 - mouse 6.83e+01  
41 43 65.2 1325 2 S16129 dynactin-associated pro 6.83e+01  
42 43 65.2 1549 2 I48103 type VII collagen - C 6.83e+01  
43 43 65.2 1735 2 A57607 Munc13-1 - rat 6.83e+01  
44 43 65.2 3473 1 A46112 genome polyprotein - 6.83e+01  
45 43 65.2 3473 2 S27927 polyprotein - rice tu 6.83e+01

## ALIGNMENTS

RESULT 1  
ENTRY BABOH #type complete  
TITLE peptide-aspartate beta-dioxygenase (EC 1.14.11.16) - bovine  
ALTERNATE\_NAMES aspartyl (asparaginyl) beta-hydroxylase  
ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle  
DATE 31-Dec-1993 #sequence\_revision 10-Feb-1995 #text\_change 29-May-1998  
ACCESSIONS A42969; A39470; B39470; C39470; S27948  
REFERENCE A42969  
#authors Jia, S.; Vandusen, W.J.; Diehl, R.E.; Kohl, N.E.; Dixon, R.A.; Elliston, K.O.; Stern, A.M.; Friedman, P.A.  
#journal J. Biol. Chem. (1992) 267:14322-14327  
#title cDNA cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase.  
#cross-references MUID:92332546  
#accession A42969  
#molecule\_type mRNA  
#residues\_type 1-754 #label JIA  
#cross-references EMBL:M91213; NID:g162693; PID:g162694  
#experimental\_source brain  
#note sequence extracted from NCBI backbone (NCBIP:108534)  
REFERENCE A39470  
#authors Wang, Q.; Vandusen, W.J.; Petroski, C.J.; Garsky, V.M.; Stern, A.M.; Friedman, P.A.  
#journal J. Biol. Chem. (1991) 266:14004-14010  
#title Bovine liver aspartyl beta-hydroxylase. Purification and characterization.  
#cross-references MUID:91310689  
#accession A39470  
#molecule\_type protein  
#residues\_type 289-328 #label WAN  
#accession B39470  
#molecule\_type protein  
#residues\_type 615,'X',617-630,'XX',633-634,'X',636,'XX',639-641  
#label WA2  
#accession C39470  
#molecule\_type protein  
#residues\_type 311-347,'X',349,'X',351-373,'X',375-379,'X',381-382  
#label WA3  
COMMENT This enzyme uses ferrous iron as a cofactor, and while beta-hydroxylating the peptidyl-aspartate substrate converts alpha-ketoglutarate to succinate and releases carbon dioxide.  
COMMENT Aspartic acid and asparagine residues in the EGF homology domain of certain plasma proteins serve as the peptidyl-aspartate substrate.  
CLASSIFICATION #superfamily peptide-aspartate beta-dioxygenase; #tetrapeptide repeat homology

```

KEYWORDS
FEATURE
2-56      #domain intracellular #status predicted #label INC\
57-78      #domain transmembrane #status predicted #label TRM\
289-754    #product peptide-aspartate beta-dioxygenase, 56K form
311-754    #status predicted #label 56K\
          #product peptide-aspartate beta-dioxygenase, 52K form
337-370    #domain tetratricopeptide repeat homology #label TTN\
371-404    #domain tetratricopeptide repeat homology #label TTA\
13,96,466,702 #binding_site carbohydrate (Asn) (covalent) #status
              predicted
SUMMARY    #length 754 #molecular-weight 84998 #checksum 9667

Query Match      98.5%; Score 65; DB 1; Length 754;
Best Local Similarity 90.0%; Pred. No. 1.90e-03;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 300 EEOQVPPET 309
      |||||
QY 1 EEOQVPPDT 10

RESULT 2
ENTRY
TITLE      #type complete
ORGANISM   aspartyl beta-hydroxylase - human
DATE      #formal_name Homo sapiens #common_name man
          29-May-1998 #sequence_revision 29-May-1998 #text_change
          10-Jul-1998
ACCESSIONS I38423
REFERENCE   I38423
#authors   Koriath, F.; Gieffers, C.; Frey, J.
#journal   Gene (1994) 150:395-399
#title     Cloning and characterization of the human gene encoding
          aspartyl beta-hydroxylase.
#cross-references MUID:95121937
#accession I38423
#status    preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues  1-757 #label RES
#cross-references EMBL:U03109; NID:9458031; PID:9458032
CLASSIFICATION #superfamily peptide-aspartate beta-dioxygenase;
               tetratricopeptide repeat homology
FEATURE
54-75      #domain transmembrane #status predicted #label TRM
SUMMARY    #length 757 #molecular-weight 85498 #checksum 2143

Query Match      98.5%; Score 65; DB 2; Length 757;
Best Local Similarity 90.0%; Pred. No. 1.90e-03;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 304 EEOQVPPET 313
      |||||
QY 1 EEOQVPPDT 10

RESULT 3
ENTRY
TITLE      #type complete
ALTERNATE_NAMES probable membrane protein YMR160w - yeast (Saccharomyces
ORGANISM   cerevisiae)
DATE      #formal_name Saccharomyces cerevisiae
          08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
          12-Dec-1997
ACCESSIONS S54518; S54605
REFERENCE   S54518
#authors   Hunt, S.; Bowman, S.
#submission submitted to the EMBL Data Library, May 1995
#accession S54518
#molecule_type DNA
#residues  1-816 #label HUN
#cross-references GB:249705; EMBL:249700; NID:9825556; PID:9825565;
               EMBL:249705; MIPS:YMR160w

```

```

GENETICS    ##experimental_source strain AB972
FEATURE
KEYWORDS    #map_position 13R
            transmembrane protein
FEATURE
328-344     #domain transmembrane #status predicted #label TMM
            #length 816 #molecular-weight 95096 #checksum 3162
SUMMARY
Query Match      83.3%; Score 55; DB 2; Length 816;
Best Local Similarity 77.8%; Pred. No. 2.96e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 24 EDOQEVPPQ 32
      |:|||||:
QY 1 EEOQEVPPD 9

RESULT 4
ENTRY
TITLE      #type complete
ORGANISM   prolactin - golden hamster
DATE      #formal_name Mesocricetus auratus #common_name golden hamster
          21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change
          16-Feb-1997
ACCESSIONS A49159
REFERENCE   A49159
#authors   Southard, J.N.; Sanchez-Jimenez, F.; Campbell, G.T.;
          Talamantes, F.
#journal   Endocrinology (1991) 129:2965-2971
#title     Sequence and expression of hamster prolactin and growth
          hormone messenger RNAs.
#cross-references MUID:92063850
#accession A49159
#status    preliminary
#molecule_type mRNA
#residues  1-226 #label SOU
#note      sequence extracted from NCBI backbone (NCBIN:66296,
          NCBIP:66298)
CLASSIFICATION #superfamily prolactin
FEATURE
33-38,85-201,
218-226      #disulfide_bonds #status predicted
SUMMARY    #length 226 #molecular-weight 25582 #checksum 3394

Query Match      74.2%; Score 49; DB 2; Length 226;
Best Local Similarity 66.7%; Pred. No. 4.98e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 97 EEOQVPPPE 105
      |||||
QY 1 EEOQEVPPD 9

RESULT 5
ENTRY
TITLE      #type complete
ORGANISM   single stranded DNA-binding protein - Aquifex aeolicus
DATE      #formal_name Aquifex aeolicus
          08-May-1998 #sequence_revision 08-May-1998 #text_change
          21-Aug-1998
ACCESSIONS H70305
REFERENCE   H70305
#authors   Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
          Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
          Keller, M.; AuJay, M.; Huber, R.; Feldman, R.A.; Short,
          J.M.; Olson, G.J.; Swanson, R.V.
#journal   Nature (1998) 392:353-358
#title     The complete genome of the hyperthermophilic bacterium
          Aquifex aeolicus.
#cross-references MUID:98196666
#accession H70305
#status    preliminary; nucleic acid sequence not shown;
          translation not shown
#molecule_type DNA
#residues  1-147 #label AQF

```

##cross-references GB:AE000672; NID:g2982810; PID:g2982816; GB:AE000657  
 ##experimental\_source strain VF5

GENETICS  
 #gene ssb  
 CLASSIFICATION #superfamily bacterial single-stranded DNA-binding protein;  
 single-stranded DNA-binding protein homology  
 SUMMARY #length 147 #molecular-weight 17132 #checksum 2119

Query Match 71.2%; Score 47; DB 2; Length 147;  
 Best Local Similarity 75.0%; Pred. No. 1.22e+01;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 115 EEEVPP 122  
 :||:||||  
 QY 1 EEQVEVPP 8

RESULT 6  
 ENTRY C22175 #type fragment  
 TITLE heat shock protein X4 - African clawed frog (fragment)  
 ORGANISM #formal\_name Xenopus laevis #common\_name African clawed frog  
 DATE 04-Sep-1998 #sequence\_revision 04-Sep-1998 #text\_change  
 24-Sep-1998

ACCESSIONS C22175  
 REFERENCE A22175  
 #authors Bienz, M.  
 #journal Proc. Natl. Acad. Sci. U.S.A. (1984) 81:3138-3142  
 #title Developmental control of the heat shock response in Xenopus.  
 #cross-references MUID:84221917  
 #accession C22175  
 #status preliminary  
 #molecule\_type mRNA  
 #residues 1-68 #label BIE

CLASSIFICATION #superfamily alpha-crystallin  
 SUMMARY #length 68 #checksum 9585  
 Query Match 69.7%; Score 46; DB 2; Length 68;  
 Best Local Similarity 55.6%; Pred. No. 1.90e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 48 DAQEPPDA 56  
 :||:||||  
 QY 2 EEQVEVPPDT 10

RESULT 7  
 ENTRY S50808 #type complete  
 TITLE hypothetical protein YJL065c - yeast (Saccharomyces  
 Cerevisiae)  
 ALTERNATE\_NAMES hypothetical protein HRD167; hypothetical protein J1115  
 ORGANISM #formal\_name Saccharomyces cerevisiae  
 DATE 13-Jan-1995 #sequence\_revision 08-Sep-1995 #text\_change  
 21-Nov-1997

ACCESSIONS S50808; S47127; S56839  
 REFERENCE S50798  
 #authors Vandenbol, M.; Durand, P.; Dion, C.; Portetelle, D.; Hilger,  
 F.

#journal Yeast (1995) 11:57-60  
 #title Sequence of a 17.1 kb DNA fragment from chromosome X of  
 Saccharomyces cerevisiae includes the mitochondrial  
 ribosomal protein L8.

#accession S50808  
 #status nucleic acid sequence not shown; translation not shown  
 #molecule\_type DNA  
 #residues 1-167 #label VAN  
 ##cross-references EMBL:234288; NID:g498992; PID:g499003  
 ##note the nucleotide sequence was submitted to the EMBL Data  
 Library, June 1994

REFERENCE S47117  
 #authors Vandenbol, M.; Durand, P.; Dion, C.; Portetelle, D.; Hilger,  
 F.  
 #submission submitted to the EMBL Data Library, June 1994

#description Sequence analysis of a 17.1 kb DNA fragment from chromosome X  
 of Saccharomyces cerevisiae includes the mitochondrial  
 ribosomal protein L8.

#accession S47127  
 #molecule\_type DNA  
 #residues 1-167 #label VAW  
 ##cross-references EMBL:234288; NID:g498992; PID:g499003

REFERENCE S56835  
 #authors Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.  
 #submission submitted to the Protein Sequence Database, September 1995  
 #accession S56839  
 #molecule\_type DNA  
 #residues 1-167 #label POH

GENETICS  
 ##cross-references EMBL:249340; NID:g1008212; PID:g1008213; MIPS:YJL065c

#map\_position 10L  
 SUMMARY #length 167 #molecular-weight 18792 #checksum 9768

Query Match 69.7%; Score 46; DB 2; Length 167;  
 Best Local Similarity 62.5%; Pred. No. 1.90e+01;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 68 QQQQVPP 75  
 :||:||||  
 QY 1 EEQVEVPP 8

RESULT 8  
 ENTRY I38346 #type fragment  
 TITLE elastic titin - human (fragment)  
 ORGANISM #formal\_name Homo sapiens #common\_name man  
 DATE 29-May-1998 #sequence\_revision 29-May-1998 #text\_change  
 05-Jun-1998

ACCESSIONS I38346  
 REFERENCE A57430  
 #authors Labelit, S.; Kolmerer, B.  
 #journal Science (1995) 270:293-296  
 #title Titins: giant proteins in charge of muscle ultrastructure and  
 elasticity.

#cross-references MUID:96026330  
 #accession I38346  
 #status preliminary; translated from GB/EMBL/DBJ  
 #molecule\_type mRNA  
 #residues 1-7962 #label RES  
 ##cross-references EMBL:X90569; NID:g1017426; PID:g1017427

GENETICS  
 #gene GDB:TTN  
 ##cross-references GDB:127867; OMIM:188840  
 #map\_position 2q31-2q31  
 SUMMARY #length 7962 #checksum 120

Query Match 69.7%; Score 46; DB 2; Length 7962;  
 Best Local Similarity 62.5%; Pred. No. 1.90e+01;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 6815 EEEVPP 6822  
 :||:||||  
 QY 2 EEQVEVPPD 9

RESULT 9  
 ENTRY JC5728 #type complete  
 TITLE neuronal UNC-119 protein - Caenorhabditis briggsae  
 ORGANISM #formal\_name Caenorhabditis briggsae  
 DATE 09-Dec-1997 #sequence\_revision 09-Dec-1997 #text\_change  
 26-Feb-1998

ACCESSIONS JC5728  
 REFERENCE JC5728  
 #authors Maduro, M.; Pilgrim, D.  
 #journal Gene (1996) 183:77-85  
 #title Conservation of function and expression of unc-119 from two  
 Caenorhabditis species despite divergence of non-coding  
 DNA.

#accession JC5728  
 ##molecule\_type DNA  
 ##residues 1-217 ##label MAD  
 ##cross-references GB:U45326; NID:g1181702; PID:g1181703  
 COMMENT This protein is involved in nervous system function.

GENETICS  
 #gene unc-119  
 #introns 13/3: 56/2; 173/1  
 SUMMARY #length 217 #molecular-weight 25094 #checksum 7052

Query Match 68.2%; Score 45; DB 2; Length 217;  
 Best Local Similarity 50.0%; Pred. No. 2.93e+01;  
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 4 EEOQSIIPGS 13  
 ||:||||:  
 QY 1 EEOQEVPPDT 10

RESULT 10  
 ENTRY S47436 #type fragment  
 TITLE flagellar antigen - Trypanosoma brucei (fragment)  
 ORGANISM #formal\_name Trypanosoma brucei  
 DATE 13-Jan-1995 #sequence\_revision 30-Jan-1998 #text\_change 30-Jan-1998

ACCESSIONS S47436  
 REFERENCE S47436  
 #authors Imboden, M.; Mueller, N.; Hemphill, A.; Mattioli, X.Y.Z.; Seebeck, T.

#submission submitted to the EMBL Data Library, August 1994  
 #description Repetitive proteins from the flagellar cytoskeleton of African Trypanosomes are diagnostically useful antigens.

#accession S47436  
 ##molecule\_type mRNA  
 ##residues 1-411 ##label IMB  
 ##cross-references EMBL:236281; NID:g530358; PID:g530359  
 ##experimental\_source strain stock TREU 1285  
 SUMMARY #length 411 #checksum 428

Query Match 68.2%; Score 45; DB 2; Length 411;  
 Best Local Similarity 50.0%; Pred. No. 2.93e+01;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 10 EEOQVPAAE 19  
 ||:||||:  
 QY 1 EEOQEVPPDT 10

RESULT 11  
 ENTRY A34373 #type complete  
 TITLE histidine-rich calcium-binding protein precursor - rabbit  
 ORGANISM #formal\_name Oryctolagus cuniculus #common\_name domestic rabbit  
 DATE 08-Jun-1990 #sequence\_revision 08-Jun-1990 #text\_change 10-Sep-1997

ACCESSIONS A34373  
 REFERENCE A34373  
 #authors Hofmann, S.L.; Goldstein, J.L.; Orth, K.; Moomaw, C.R.; Slaughter, C.A.; Brown, M.S.

#journal J. Biol. Chem. (1989) 264:18083-18090  
 #title Molecular cloning of a histidine-rich Ca(2+)-binding protein of sarcoplasmic reticulum that contains highly conserved repeated elements.

#cross-references MUID:90036884

#accession A34373  
 ##status preliminary  
 ##molecule\_type mRNA  
 ##residues 1-852 ##label HOF  
 ##cross-references GB:J05080; NID:g165099; PID:g165100

KEYWORDS calcium binding

SUMMARY #length 852 #molecular-weight 96116 #checksum 3434

Query Match 68.2%; Score 45; DB 2; Length 852;

Best Local Similarity 55.6%; Pred. No. 2.93e+01;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 190 EEEEVSPSE 198  
 ||:||||:  
 QY 1 EEOQEVPPD 9

RESULT 12  
 ENTRY A46068 #type complete  
 TITLE Huntingdon disease-associated protein - human  
 ORGANISM #formal\_name Homo sapiens #common\_name man  
 DATE 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 29-Aug-1997

ACCESSIONS A46068; I54337  
 REFERENCE A46068  
 #authors MacDonald, M.E.; Ambrose, C.M.; Duyao, M.P.; Myers, R.H.; Lin, C.; Srinidhi, L.; Barnes, G.; Taylor, S.A.; James, M.; Groot, N.; MacFarlane, H.; Jenkins, B.; Anderson, M.A.; Wexler, N.S.; Gusella, J.F.; Bates, G.P.; Baxendale, S.; Hummerich, H.; Kirby, S.; North, M.; Youngman, S.; Mott, R.; Zehetner, G.; Sedlacek, Z.; Poustka, A.; Frischauf, A.M.; Buckler, A.J.; Church, D.; Doucette-Stamm, L.; O'Donovan, M.C.; Riba-Ramirez, L.; Shah, M.; Stanton, V.P.; Strobel, S.A.; Draths, K.M.; Wales, J.L.; Dervan, P.; Housman, D.E.; Altherr, M.; Shiang, R.; Thompson, L.; Fielder, T.; Wasmuth, J.J.; Tagle, D.; Valdes, J.; Elmer, L.; Allard, M.; Castilla, L.; Swaroop, M.; Blanchard, K.; Collins, F.S.; Snell, R.; Holloway, T.; Gillespie, K.; Datson, N.; Shaw, D.; Harper, P.S.

#journal Cell (1993) 72:971-983  
 #title A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes.  
 #accession A46068  
 ##status preliminary  
 ##molecule\_type mRNA  
 ##residues 1-3144 ##label MAC  
 ##cross-references GB:I12392  
 REFERENCE I54337  
 #authors Lin, B.; Rommens, J.M.; Graham, R.K.; Kalchman, M.; Hayden, M.R.; Macdonald, H.; Nasir, J.; Delaney, A.; Goldberg, Y.P.; Hayden, M.R.

#journal Hum. Mol. Genet. (1993) 2:1541-1545  
 #title Differential 3' polyadenylation of the Huntington disease gene results in two mRNA species with variable tissue expression.  
 #cross-references MUID:94093536  
 #accession I54337  
 ##status preliminary; translated from GB/EMBL/DBJ  
 ##molecule\_type mRNA  
 ##residues 2563-3144 ##label RES  
 ##cross-references GB:L20431; NID:g398028; PID:g398029

GENETICS  
 #gene GDB:HD  
 #cross-references GDB:119307; OMIM:143100  
 #map\_position 4p16.3-4p16.3  
 SUMMARY #length 3144 #molecular-weight 347896 #checksum 1801

Query Match 68.2%; Score 45; DB 2; Length 3144;  
 Best Local Similarity 60.0%; Pred. No. 2.93e+01;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 EEEEDVPNT 2352  
 ||:||||:  
 QY 1 EEOQEVPPDT 10

RESULT 13  
 ENTRY S37431 #type complete

TITLE ankyrin 2, neuronal long splice form - human  
 ALTERNATE\_NAMES ankyrin B, 440K splice form; ankyrin-B; brain ankyrin; non-erythroid ankyrin  
 CONTAINS ankyrin 2, short form

```

430-462      #domain ankyrin repeat homology #label AN12\
463-495      #domain ankyrin repeat homology #label AN13\
496-528      #domain ankyrin repeat homology #label AN14\
529-561      #domain ankyrin repeat homology #label AN15\
562-594      #domain ankyrin repeat homology #label AN16\
595-627      #domain ankyrin repeat homology #label AN17\
628-660      #domain ankyrin repeat homology #label AN18\
661-693      #domain ankyrin repeat homology #label AN19\
694-726      #domain ankyrin repeat homology #label AN20\
727-759      #domain ankyrin repeat homology #label AN21\
760-792      #domain ankyrin repeat homology #label AN22\
793-825      #domain ankyrin repeat homology #label AN23\
SUMMARY      #length 3924 #molecular-weight 430340 #checksum 3664

Query Match      68.2%; Score 45; DB 2; Length 3924;
Best Local Similarity 40.0%; Pred. No. 2.93e+01;
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 3839 DDMPEIPPT 3848
Qy 1 EEQEVPPD 10
:: |::|:|

RESULT 14
ENTRY TITLE
S37862 #type complete
hypothetical protein YKL041w - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein YKL254
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 03-May-1994 #sequence_revision 03-May-1994 #text_change 06-Feb-1998
ACCESSIONS S37862; S40650
REFERENCE S37851
#authors Purnelle, B.; Skala, J.; van Dyck, L.; Tettelin, H.; Goffeau, A.
#submission submitted to the Protein Sequence Database, March 1994
#accession S37862
##molecule_type DNA
##residues 1-224 ##label PUR
#cross-references EMBL:J28041; NID:g486051; PID:g486052; MIPS:YKL041w
#experimental_source strain S288C
REFERENCE S40650
#authors Purnelle, B.; Tettelin, H.; van Dyck, L.; Skala, J.; Goffeau, A.
#journal Yeast (1993) 9:1379-1384
#title The sequence of a 17.5 kb DNA fragment on the left arm of yeast chromosome XI identifies the protein kinase gene ELM1, the DNA primase gene PR12, a new gene encoding a putative histone and seven new open reading frames.
#accession S40650
#status nucleic acid sequence not shown
##molecule_type DNA
##residues 1-224 ##label PUR2
#cross-references EMBL:X71621; NID:g1478265; PID:g666099
#experimental_source strain S288C
GENETICS
#gene SGD:VPS24
#map_position 11L
#cross-references SGD:S0001524; MIPS:YKL041w
SUMMARY #length 224 #molecular-weight 26242 #checksum 5062

Query Match      66.7%; Score 44; DB 2; Length 224;
Best Local Similarity 55.6%; Pred. No. 4.48e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 194 EEEQIPDE 202
Qy 1 EEQEVPPD 9
|::|::|:|

RESULT 15
ENTRY TITLE
H65206 #type complete
urocorbryinogen decarboxylase (EC 4.1.1.37) - Escherichia

```

```

coli
#formal_name Escherichia coli
12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
05-Dec-1997
ACCESSIONS H65206; JN0894; JS0708
REFERENCE A64720
#authors
Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal
Science (1997) 277:1453-1462
#title
The complete genome sequence of Escherichia coli K-12.
#cross-references MIM:97426617
#accession H65206
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-354 #label BLAT
##cross-references GB:A5000473; GB:U00096; NID:g2367336; PID:g2367337;
UWGP:b3997
##experimental_source strain K-12, substrain MG1655
REFERENCE JN0894
#authors
Nishimura, K.; Nakayashiki, T.; Inokuchi, H.
#journal
Gene (1993) 133:109-113
#title
Cloning and sequencing of the hemE gene encoding
uroporphyrinogen III decarboxylase (UFD) from Escherichia
coli K-12.
#accession JN0894
##status nucleic acid sequence not shown
##molecule_type DNA
##residues 1-71,'RSS','75','R','Y','79-82','I','84-88','SSILKP','95','KV',
98,'VLPRQI','104-250','SAIV','A','257-354 #label N12
##cross-references GB:D12624; NID:g216563; PID:d1002638; PID:g216564
##experimental_source strain K-12
COMMENT This enzyme catalyzes the decarboxylation of uroporphyrinogen III
to coproporphyrinogen.
GENETICS
#gene hemeF
#map_position 90 min
CLASSIFICATION #superfamily uroporphyrinogen decarboxylase
KEYWORDS carbon-carbon lyase; carboxy-lyase; lyase; porphyrin
biosynthesis
SUMMARY #length 354 #molecular-weight 39248 #checksum 7044
Query Match 66.7%; Score 44; DB 2; Length 354;
Best Local Similarity 57.1%; Pred. No. 4.48e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 330 HODVPP 336
QY :||||:
3 QDEVPPD 9
Search completed: Thu May 13 15:32:28 1999
Job time : 14 secs.

```

\*\*\*\*\*

MAESREH (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu May 13 15:32:47 1999; MasPar time 2.01 Seconds  
Tabular output not generated. 133.387 Million cell updates/sec

Title: >US-09-040-485-7  
Description: (1-10) from US09040485.pap  
Perfect Score: 66  
Sequence: 1 EEQQEVPDPDT 10

Scoring table: PAM 150  
Gap 15

Searched: 74019 seqs, 26840295 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot35  
1:swissprot

Statistics: Mean 22.125; Variance 25.762; scale 0.859

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID           | Description            | Pred. No. |
|------------|-------|-------------|--------|--------------|------------------------|-----------|
| 1          | 65    | 98.5        | 754    | 1 ASPH_BOVIN | ASPARTYL/ASPARAGINYL B | 2.67e-04  |
| 2          | 65    | 98.5        | 757    | 1 ASPH_HUMAN | ASPARTYL/ASPARAGINYL B | 2.67e-04  |
| 3          | 55    | 83.3        | 816    | 1 YM35_YEAST | HYPOTHETICAL 95.1 KD P | 7.70e-02  |
| 4          | 49    | 74.2        | 226    | 1 PRL_MESAU  | PROLACTIN PRECURSOR (P | 1.79e+00  |
| 5          | 48    | 72.7        | 815    | 1 LUI5_HUMAN | PUTATIVE TUMOR SUPPRES | 2.95e+00  |
| 6          | 46    | 69.7        | 167    | 1 YJG5_YEAST | HYPOTHETICAL 18.8 KD P | 7.87e+00  |
| 7          | 46    | 69.7        | 775    | 1 LY54_EMENI | HOMOACONITASE PRECURS  | 7.87e+00  |
| 8          | 45    | 68.2        | 217    | 1 PEXH_YARLI | UNC-119 PROTEIN.       | 1.27e+01  |
| 9          | 45    | 68.2        | 671    | 1 SRCH_RABIT | PEROXISOMAL MEMBRANE P | 1.27e+01  |
| 10         | 45    | 68.2        | 852    | 1 ANKH_HUMAN | SARCOPLASMIC RETICULUM | 1.27e+01  |
| 11         | 45    | 68.2        | 1839   | 1 HD_HUMAN   | ANKYRIN, BRAIN VARIANT | 1.27e+01  |
| 12         | 45    | 68.2        | 3144   | 1 ANKH_HUMAN | HUNTINGTIN (HUNTINGTON | 1.27e+01  |
| 13         | 45    | 68.2        | 3924   | 1 ANKH_HUMAN | ANKYRIN, BRAIN VARIANT | 1.27e+01  |
| 14         | 44    | 66.7        | 224    | 1 YKEL_YEAST | HYPOTHETICAL 26.2 KD P | 2.04e+01  |
| 15         | 44    | 66.7        | 354    | 1 DCUP_ECOLI | UROPORPHYRINOGEN DECAR | 2.04e+01  |
| 16         | 44    | 66.7        | 414    | 1 NOP3_HUMAN | NUCLEOLAR PROTEIN 3 (M | 2.04e+01  |
| 17         | 44    | 66.7        | 700    | 1 MYBB_HUMAN | MYB-RELATED PROTEIN B  | 2.04e+01  |
| 18         | 44    | 66.7        | 755    | 1 RREL_HUMAN | RAS-RESPONSIVE ELEMENT | 2.04e+01  |
| 19         | 43    | 65.2        | 140    | 1 ANF_CHICK  | ATRIAL NATRIURETIC FAC | 3.24e+01  |
| 20         | 43    | 65.2        | 184    | 1 K501_ACTH  | FRUIT PROTEIN PKIWI501 | 3.24e+01  |
| 21         | 43    | 65.2        | 395    | 1 CG2A_CHICK | G2/MITOTIC-SPECIFIC CY | 3.24e+01  |
| 22         | 43    | 65.2        | 404    | 1 ARRS_BOVIN | S-ARRESTIN (RETINAL S- | 3.24e+01  |
| 23         | 43    | 65.2        | 405    | 1 ARRS_HUMAN | S-ARRESTIN (RETINAL S- | 3.24e+01  |

ALIGNMENTS

| RESULT | 1  | ASPH_BOVIN  | STANDARD; | PRT; | 754 AA. |
|--------|--|---|-----------|------|---------|
| ID     | AC   | Q28056;   |           |      |         |
| DT     | 01-NOV-1997  | (REL. 35, CREATED)  |           |      |         |
| DT     | 01-NOV-1997  | (REL. 35, LAST SEQUENCE UPDATE)                             |           |      |         |
| DT     | 01-NOV-1997  | (REL. 35, LAST ANNOTATION UPDATE)                           |           |      |         |
| DE     | ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-HYDROXYLASE)   | (ASP BETA-HYDROXYLASE) (PEPTIDE-ASPARTATE BETA-HYDROXYLASE) |           |      |         |
| GN     | ASPH.  |   |           |      |         |
| OS     | BOS TAURUS (BOVINE).   |   |           |      |         |
| OC     | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;   |   |           |      |         |
| OC     | EUTHERIA; ARTIODACTYLA.  |   |           |      |         |
| RN     | [1]  |   |           |      |         |
| RC     | SEQUENCE FROM N.A.   |   |           |      |         |
| RP     | TISSUE=LIVER, AND BRAIN;   |   |           |      |         |
| RX     | MEDLINE: 92332546.   |   |           |      |         |
| RA     | JIA S., VANDUSEN W.J., DIEHL R.E., KOHL N.E., DIXON R.A.F.,  |   |           |      |         |
| RA     | ELLISTON K.O., STERN A.M., FRIEDMAN P.A.;  |   |           |      |         |
| RL     | J. BIOL. CHEM. 267:14322-14327(1992).  |   |           |      |         |
| RN     | [2]  |   |           |      |         |
| RP     | SEQUENCE OF 289-385 AND 615-641.   |   |           |      |         |
| RC     | TISSUE=LIVER;  |   |           |      |         |
| RX     | MEDLINE: 91310689.   |   |           |      |         |
| RA     | WANG Q., VANDUSEN W.J., PETROSKI C.J., GARSKY V.M., STERN A.M.,  |   |           |      |         |
| RA     | FRIEDMAN P.A.;   |   |           |      |         |
| RL     | J. BIOL. CHEM. 266:14004-14010(1991).  |   |           |      |         |
| CC     | - FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF PROTEINS. |   |           |      |         |
| CC     | - CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) -  |   |           |      |         |
| CC     | PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).   |   |           |      |         |
| CC     | - COFACTOR: IRON.  |   |           |      |         |
| CC     | - SUBUNIT: MONOMER.  |   |           |      |         |
| CC     | - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM.   |   |           |      |         |
| CC     | - PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 289-754) OR 52 KD (AA 311-754) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM.                 |   |           |      |         |
| DR     | EMBL: M91213; G162694;   |   |           |      |         |
| KW     | OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR; ENDOPLASMIC RETICULUM.  |   |           |      |         |
| FT     | DOMAIN 1 57  |   |           |      |         |
| FT     | TRANSMEM 58 78   |   |           |      |         |
| FT     | CYTOPLASMIC (POTENTIAL).   |   |           |      |         |
| FT     | SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)   |   |           |      |         |
| FT     | (POTENTIAL)  |   |           |      |         |
| FT     | LUMENAL (POTENTIAL).   |   |           |      |         |
| FT     | POLY-GLY.  |   |           |      |         |
| FT     | POLY-SER.  |   |           |      |         |

```

FT DOMAIN 318 328 POLY-LYS.
FT CARBOHYD 96 96 POTENTIAL.
FT CARBOHYD 466 466 POTENTIAL.
FT CARBOHYD 702 702 POTENTIAL.
SQ SEQUENCE 754 AA; 84998 MW; 608861B2 CRC32;

Query Match 98.5%; Score 65; DB 1; Length 754;
Best Local Similarity 90.0%; Pred. No. 2.67e+04;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 300 EEOQEVPPET 309
QY 1 EEOQEVPPDT 10

RESULT 2
ID ASPH_HUMAN STANDARD; PRT; 757 AA.
AC Q12797;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.4.11.16) (ASPARTATE BETA-
DE HYDROXYLASE) (ASP BETA-HYDROXYLASE) (PEPTIDE-ASPARTATE BETA-
DE DIOXYGENASE).
GN ASPH.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 95121937.
RA KORIOTH F., GIEFFERS C., FREY J.;
RL GENE 150:395-399(1994).
CC -!- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN
CC CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF
CC PROTEINS.
CC -!- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) -
CC PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).
CC -!- COFACTOR: IRON.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC
CC RETICULUM.
CC -!- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES TESTED.
CC -!- PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 274-757) OR 52 KD (AA
CC 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
DR EMBL; U03109; E82591; -.
DR MIM; 600582; -.
KW OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR;
KW ENDOPLASMIC RETICULUM.
FT DOMAIN 1 54 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 55 75 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 76 757 LUMENAL (POTENTIAL).
FT DOMAIN 13 20 POLY-SER.
FT DOMAIN 323 332 POLY-LYS.
FT CARBOHYD 452 452 POTENTIAL.
FT CARBOHYD 705 705 POTENTIAL.
SQ SEQUENCE 757 AA; 85498 MW; AE6AFC24 CRC32;

Query Match 98.5%; Score 65; DB 1; Length 757;
Best Local Similarity 90.0%; Pred. No. 2.67e+04;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 304 EEOQEVPPET 313
QY 1 EEOQEVPPDT 10

RESULT 3
ID YM35_YEAST STANDARD; PRT; 816 AA.
AC Q03823;
DT 01-NOV-1997 (REL. 35, CREATED)

```

```

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 95.1 KD PROTEIN IN IMPI-HLJ1 INTERGENIC REGION.
GN YMR160W OR YMR520.09.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA HUNT S., BOWMAN S., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; Z49705; G825565; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 816 AA; 95096 MW; A897E595 CRC32;

Query Match 83.3%; Score 55; DB 1; Length 816;
Best Local Similarity 77.8%; Pred. No. 7.70e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 24 EEOQEVPPQ 32
QY 1 EEOQEVPPD 9

RESULT 4
ID PRL_MESAU STANDARD; PRT; 226 AA.
AC P37884;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE PROLACTIN PRECURSOR (PRL).
GN PRL.
OS MESOCRICETUS AURATUS (GOLDEN HAMSTER).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN SEQUENCE FROM N.A.
RX MEDLINE; 92063850.
RA SOUTHWARD J.N., SANCHEZ-JIMENEZ F., CAMPBELL G.T., TALAMANTES F.;
RL ENDOCRINOLOGY 129:2965-2971(1991).
CC -!- FUNCTION: PROLACTIN ACTS PRIMARILY ON THE MAMMARY GLAND BY
CC PROMOTING LACTATION.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
DR EMBL; S86296; G239354; -.
DR PIR; A49159; A49159.
DR PROSITE; PS00266; SOMATOTROPIN_1; 1.
DR PROSITE; PS00338; SOMATOTROPIN_2; 1.
KW HORMONE; PARTURITION; LACTATION; PITUITARY; SIGNAL.
FT CHAIN 1 29 BY SIMILARITY.
FT SIGNAL 30 226 PROLACTIN.
FT DISULFID 33 38 BY SIMILARITY.
FT DISULFID 85 201 BY SIMILARITY.
FT DISULFID 218 226 BY SIMILARITY.
SQ SEQUENCE 226 AA; 25582 MW; D5A4C5BC CRC32;

Query Match 74.2%; Score 49; DB 1; Length 226;
Best Local Similarity 66.7%; Pred. No. 1.79e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 97 EEOQVPPPE 105
QY 1 EEOQEVPPD 9

RESULT 5
ID LU15_HUMAN STANDARD; PRT; 815 AA.
AC P52756;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PUTATIVE TUMOR SUPPRESSOR LOC15.
GN LOC15.

```



OC EUNARIOIA; FUNGI; ASCOMICOTINA; HEMIASCOMICETI

OC EUNARIOIA; FUNGI; ASCOMICOTINA; HEMIASCOMICETI

RN SEQUENCE FROM N.A.  
 RP STRAIN-E122;  
 RC MEDLINE; 97265383.  
 RA SMITH J.J., SZILARD R.K., MARELLI M., RACHUBINSKI R.A.;  
 RL MOL. CELL. BIOL. 17:2511-2520(1997).  
 CC -!- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS. REQUIRED FOR THE  
 CC IMPORT OF A SUBSET OF MATRIX PROTEINS.  
 CC -!- SUBCELLULAR LOCATION: PEROXISOMAL MEMBRANE-ASSOCIATED.  
 DR EMBL; U73028; G2039277; -.  
 KW PEROXISOME; TRANSMEMBRANE.  
 FT TRANSMEM 127 147 . POTENTIAL.  
 FT TRANSMEM 187 207 POTENTIAL.  
 FT TRANSMEM 212 232 POTENTIAL.  
 FT TRANSMEM 258 278 POTENTIAL.  
 FT TRANSMEM 318 338 POTENTIAL.  
 FT TRANSMEM 372 392 POTENTIAL.  
 FT TRANSMEM 475 495 POTENTIAL.  
 FT TRANSMEM 504 524 POTENTIAL.  
 SQ SEQUENCE 671 AA; 75588 MW; BC5618B9 CRC32;

Query Match 68.2%; Score 45; DB 1; Length 671;  
 Best Local Similarity 50.0%; Pred. No. 1.27e+01;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 573 OEQEDIPP 580  
 QY :||:||||  
 1 EEQEVVPP 8

RESULT 10  
 ID SRCH RABIT STANDARD; PRT; 852 AA.  
 AC P16230;  
 DT 01-APR-1990 (REL. 14, CREATED)  
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
 DE SARCOPLASMIC RETICULUM HISTIDINE-RICH CALCIUM-BINDING PROTEIN  
 DE PRECURSOR (HCP).  
 OS ORYCTOLAGUS CUNICULUS (RABBIT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; LAGOMORPHA.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE-SKELETAL MUSCLE;  
 RX MEDLINE; 90036884.  
 RA HOFMANN S.L., GOLDSTEIN J.L., ORTH K., MOOMAW C.R., SLAUGHTER C.A.,  
 RA BROWN M.S.;  
 RL J. BIOL. CHEM. 264:18083-18090(1989).  
 CC -!- FUNCTION: HCP MAY PLAY A ROLE IN THE REGULATION OF CA(2+)  
 CC SEQUESTRATION OR RELEASE IN THE SR OF SKELETAL AND CARDIAC  
 CC MUSCLE.  
 CC -!- SUBCELLULAR LOCATION: SARCOPLASMIC RETICULUM LUMEN.  
 CC EMBL; J05080; G165100; -.  
 DR PIR; A34373; A34373.  
 DR PROSITE; PS00328; HCP; 10.  
 KW CALCIUM-BINDING; SIGNAL; REPEAT.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT CHAIN 28 852 HISTIDINE-RICH CALCIUM-BINDING PROTEIN.  
 FT MOD\_RES 28 28 BLOCKED.  
 FT DOMAIN 59 100 2 X APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 59 79 1-1.  
 FT REPEAT 80 100 1-2.  
 FT DOMAIN 199 470 10 X TANDEM REPEATS, ACIDIC.  
 FT REPEAT 199 224 2-1.  
 FT REPEAT 225 233 2-2.  
 FT REPEAT 254 282 2-3.  
 FT REPEAT 283 310 2-4.  
 FT REPEAT 311 339 2-5.  
 FT REPEAT 340 367 2-6.  
 FT REPEAT 368 395 2-7.  
 FT REPEAT 396 423 2-8.  
 FT REPEAT 424 451 2-9.

FT REPEAT 452 470 2-10.  
 FT DOMAIN 471 585 4 X APPROXIMATE TANDEM REPEATS.  
 FT DOMAIN 721 733 POLY-GLU.  
 FT DOMAIN 780 826 METAL-BINDING (POTENTIAL).  
 SQ SEQUENCE 852 AA; 96117 MW; 3C1DF781 CRC32;  
 Query Match 68.2%; Score 45; DB 1; Length 852;  
 Best Local Similarity 55.6%; Pred. No. 1.27e+01;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Db 190 EEEVEVSPE 198  
 QY ||:||||  
 1 EEQEVVPP 9  
 RESULT 11  
 ID ANKC\_HUMAN STANDARD; PRT; 1839 AA.  
 AC Q01485;  
 DT 01-APR-1993 (REL. 25, CREATED)  
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE ANKYRIN, BRAIN VARIANT 2 (ANKYRIN B) (ANKYRIN, NONERYTHROID)  
 DE (FRAGMENT).  
 GN ANK2.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN STEM;  
 RX MEDLINE; 91302466.  
 RA OTTO E., KUNIMOTO M., MCLAUGHLIN T., BENNETT V.;  
 RL J. CELL BIOL. 114:241-253(1991).  
 RN [2]  
 RP SEQUENCE OF 463-495 FROM N.A.  
 RX MEDLINE; 92009921.  
 RA TSE W.T., MENNINGER J.C., YANG-FENG T.L., FRANCKE U., SAHR K.E.,  
 RA LUX S.E., WARD D.C., FORGET B.G.;  
 RA GENOMICS 10:858-866(1991).  
 CC -!- FUNCTION: ANKYRINS ATTACH INTEGRAL MEMBRANE PROTEINS TO CYTO-  
 CC SKELETAL ELEMENTS; THEY BIND TO THE ERYTHROCYTE MEMBRANE PROTEIN  
 CC BAND 4.2, TO NA-K ATPASE, TO THE LYMPOCYTE MEMBRANE PROTEIN GP85,  
 CC AND TO THE CYTOSKELETAL PROTEINS FODRIN, TUBULIN, VIMENTIN AND  
 CC DESMIN. ERYTHROCYTE ANKYRINS ALSO LINK SPECTRIN (BETA CHAIN) TO  
 CC THE CYTOPLASMIC DOMAIN OF THE ERYTHROCYTES ANION EXCHANGE PROTEIN;  
 CC THEY RETAIN MOST OR ALL OF THESE BINDING FUNCTIONS.  
 CC -!- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL  
 CC CELLS THROUGHOUT THE BRAIN.  
 CC -!- ALTERNATIVE PRODUCTS: THE TWO BRAIN VARIANTS ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING OF GENE ANK2.  
 CC -!- SIMILARITY: CONTAINS 24 ANK REPEATS.  
 DR EMBL; X56958; G29491; -.  
 DR EMBL; M37123; G178648; -.  
 DR PIR; S14569; S14569.  
 DR PIR; B39643; B39643.  
 DR MIM; 106410; -.  
 DR PROSITE; PS50017; DEATH\_DOMAIN; 1.  
 KW CYTOSKELETON; ALTERNATIVE SPLICING; REPEAT; ANK REPEAT;  
 KW MULTIGENE FAMILY.  
 FT DOMAIN 63 792 22 X ANK MOTIF REPEATS.  
 FT REPEAT 63 95 ANK MOTIF 1.  
 FT REPEAT 96 128 ANK MOTIF 2.  
 FT REPEAT 129 161 ANK MOTIF 3.  
 FT REPEAT 162 190 ANK MOTIF 4.  
 FT REPEAT 191 231 ANK MOTIF 5.  
 FT REPEAT 232 264 ANK MOTIF 6.  
 FT REPEAT 265 297 ANK MOTIF 7.  
 FT REPEAT 298 330 ANK MOTIF 8.  
 FT REPEAT 331 363 ANK MOTIF 9.  
 FT REPEAT 364 396 ANK MOTIF 10.  
 FT REPEAT 397 429 ANK MOTIF 11.  
 FT REPEAT 430 462 ANK MOTIF 12.  
 FT REPEAT 463 495 ANK MOTIF 13.

```

FT REPEAT 496 528 ANK MOTIF 14.
FT REPEAT 529 561 ANK MOTIF 15.
FT REPEAT 562 594 ANK MOTIF 16.
FT REPEAT 595 627 ANK MOTIF 17.
FT REPEAT 628 660 ANK MOTIF 18.
FT REPEAT 661 693 ANK MOTIF 19.
FT REPEAT 694 726 ANK MOTIF 20.
FT REPEAT 727 759 ANK MOTIF 21.
FT REPEAT 760 792 ANK MOTIF 22.
FT DOMAIN 1451 1535 DEATH DOMAIN.
FT CONFLICT 475 476 GQ -> PE (IN REF. 2).
FT NON_TER 1839 1839
SQ SEQUENCE 1839 AA; 202409 MW; 546A50B4 CRC32;

Query Match 68.2%; Score 45; DB 1; Length 1839;
Best Local Similarity 40.0%; Pred. No. 1.27e+01;
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 1754 DMPPIPET 1763
Qy 1 EEQVEPPT 10

RESULT 12
ID HD_HUMAN STANDARD; PRT; 3144 AA.
AC P42858;
DT 01-NOV-1995 (REL. 32; CREATED)
DT 01-NOV-1995 (REL. 32; LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36; LAST ANNOTATION UPDATE)
DE HUNTINGTIN (HUNTINGTON'S DISEASE PROTEIN) (HD PROTEIN).
GN HD OR IT15.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RX MEDLINE; 93208892.
RA MACDONALD M., AMBROSE C.M., DUYAO M.P., MYERS R.H., LIN C.S.,
RA SRINIDHI J., BARNES G., TAYLOR S.A., JAMES M., GROOT N., MCFARLANE H.,
RA JENKINS B., ANDERSON M.A., WEXLER N.S., GUSELLA J.F., BATES G.P.,
RA BAENDALE S., HUMMERICH H., KIRBY S., NORTH M., YOUNGMAN S., MOTT R.,
RA ZEHNER G., SEDLACEK Z., POUSTKA A., FRISCHAUF A.-M., LEHRACH H.,
RA BUCKLER A.J., CHURCH D., DOUCETTE-STAMM L., O'DONOVAN M.C.,
RA RIBA-RAMIREZ L., SHAH M., STANTON V.P., STROBEL S.A., DRATHS K.M.,
RA WALES J.L., DERVAN P., HOUSMAN D.E., ALTHERR M., SHANG R.,
RA THOMPSON L., FIELDER T., WASMUTH J.J., TAGLE D., VALDES J., ELMER L.,
RA ALLARD M., CASTILLA L., SWAROOP M., BLANCHARD K., COLLINS F.S.,
RA SNELL R., HOLLOWAY T., GILLESPIE K., DATSON N., SHAW S., HARPER P.S.;
RA CELL 72:971-983(1993).
RN [2]
RP SEQUENCE OF 1-90 FROM N.A.
RX MEDLINE; 95278941.
RA LIN B., NASIR J., KALCHMAN M.A., MCDONALD H., ZEISLER J.,
RA GOLDBERG Y.P., HAYDEN M.R.;
RA GENOMICS 25:707-715(1995).
RN [3]
RP SEQUENCE OF 1-205 FROM N.A.
RX MEDLINE; 94255787.
RA AMBROSE C.M., DUYAO M.P., BARNES G., BATES G.P., LIN C.S.,
RA SRINIDHI J., BAENDALE S., HUMMERICH H., LEHRACH H., ALTHERR M.,
RA WASMUTH J., BUCKLER A., CHURCH D., HOUSMAN D., BERKS M., MICKLEM G.,
RA DUBURN R., DODGE A., READ A., GUSELLA J.F., MACDONALD M.E.;
RA SOMAT. CELL MOL. GENET. 20:27-38(1994).
RN [4]
RP SEQUENCE OF 1-117 FROM N.A.
RX MATTHEWS P.;
RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [5]
RP SEQUENCE OF 119-934 FROM N.A.
RA LLOYD C.;
RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [6]

```

```

RP SEQUENCE OF 1212-1290 FROM N.A.
RA MUNGALL A., ODELL C.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [7]
RP SEQUENCE OF 1291-1860 FROM N.A.
RA MUNGALL A.;
RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [8]
RP SEQUENCE OF 1862-2820 FROM N.A.
RA BUCK D.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [9]
RP SEQUENCE OF 2563-3144 FROM N.A.
RC TISSUE=FRONTAL CORTEX, BRAIN, RETINA, CAUDATE, AND MUSCLE;
RX MEDLINE; 94093536.
RA LIN B., ROMMENS J.M., GRAHAM R.K., KALCHMAN M., MACDONALD H.,
RA NASIR J., DELANEY A., GOLDBERG Y.P., HAYDEN M.R.;
RA HUM. MOL. GENET. 2:1341-1545(1993).
RN [10]
RP SUBCELLULAR LOCATION.
RX MEDLINE; 95375771.
RA TROTIER Y., DEVIS D., IMBERT G., SAUDOU F., AN I., LUTZ Y., WEBER C.,
RA ACID Y., HIRSCH E.C., MANDEL J.-L.;
RA NAT. GENET. 10:104-110(1995).
RN [11]
RP CLEAVAGE BY APOPAIN.
RX MEDLINE; 96331285.
RA GOLDBERG Y.P., NICHOLSON D.W., RASPER D.M., KALCHMAN M.A., KOIDE H.B.,
RA GRAHAM R.K., BROMM M., KAZEMI-ESFARJANI P., THORNBERRY N.A.,
RA VAILLANCOURT J.P., HAYDEN M.R.;
RA NAT. GENET. 13:442-449(1996).
CC -!- FUNCTION: MAY PLAY A ROLE IN MICROTUBULE-MEDIATED TRANSPORT OR
CC VESICLE FUNCTION.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED IN A VARIETY OF TISSUES WITH
CC THE HIGHEST LEVEL EXPRESSION IN THE BRAIN (NERVE FIBRES,
CC VARICOSITIES, AND NERVE ENDINGS). IN THE BRAIN, THE REGIONS WHERE
CC IT CAN BE MAINLY FOUND ARE THE CEREBELLAR CORTEX, THE NEOCORTX,
CC THE STRIATUM, AND THE HIPPOCAMPAL FORMATION.
CC -!- PTM: CLEAVED BY APOPAIN DOWNSTREAM OF THE POLYGLUTAMINE STRETCH.
CC THE RESULTING AMINO-TERMINAL FRAGMENT IS CYTOTOXIC AND PROVOKES
CC APOPTOSIS.
CC -!- POLYMORPHISM: THE POLY-GLN REGION OF HD IS HIGHLY POLYMORPHIC (10
CC TO 35 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO ABOUT
CC 36-120 REPEATS IN HD PATIENTS. THE REPEAT LENGTH USUALLY INCREASES
CC IN SUCCESSIVE GENERATIONS, BUT CONTRACTS ALSO ON OCCASION. THE
CC LONGER EXPANSIONS RESULT IN EARLIER ONSET AND MORE SEVERE CLINICAL
CC MANIFESTATIONS OF THE DISEASE. THE ADJACENT POLY-PRO REGION IS
CC ALSO POLYMORPHIC AND VARIES BETWEEN 7-12 RESIDUES. POLYGLUTAMINE
CC EXPANSION LEADS TO ELEVATED SUSCEPTIBILITY TO APOPAIN CLEAVAGE AND
CC LIKELY RESULT IN ACCELERATED NEURONAL APOPTOSIS.
CC -!- DISEASE: DEFECTS IN HD ARE THE CAUSE OF HUNTINGTON'S DISEASE, AN
CC AUTOSOMAL DOMINANT NEURODEGENERATIVE DISORDER CHARACTERIZED BY
CC INVOLUNTARY MOVEMENTS (CHOREA), GENERAL MOTOR IMPAIRMENT,
CC PSYCHIATRIC DISORDERS AND DEMENTIA. ONSET OF THE DISEASE OCCURS
CC USUALLY IN THE THIRD OR FOURTH DECADE OF LIFE AND SYMPTOMS
CC PROGRESSIVELY WORSEN LEADING TO DEATH IN 10 TO 20 YEARS. IT
CC AFFECTS 1 IN 10,000 INDIVIDUALS OF EUROPEAN ORIGIN. NEUROPATHOLOGY
CC OF HUNTINGTON'S DISEASE DISPLAYS A DISTINCTIVE PATTERN WITH LOSS
CC OF NEURONS, SPECIALLY IN THE CAUDATE AND PUTAMEN (STRIATUM).
CC -!- SIMILARITY: STRONGLY CONSERVED BETWEEN HUMAN, RAT AND MOUSE.
CC -!- DATABASE: NAME=HotMolbase; NOTE=HD entry;
CC WWW="http://bioinformatics.weizmann.ac.il/hotmolbase/entries/hunt1.htm".
CC EMBL; L12392; G454415;
CC EMBL; L34020; -; NOT_ANNOTATED_CDS.
CC EMBL; L27350; -; NOT_ANNOTATED_CDS.
CC EMBL; L27351; -; NOT_ANNOTATED_CDS.
CC EMBL; L27352; -; NOT_ANNOTATED_CDS.
CC EMBL; L27353; -; NOT_ANNOTATED_CDS.
CC EMBL; L27354; -; NOT_ANNOTATED_CDS.
CC EMBL; L27355; E218669; -;
CC EMBL; L249155; G794068; -;
CC EMBL; L249208; G798839; -;

```

DR EMBL; 269649; E223397; -  
 DR EMBL; 249154; G794066; -  
 DR EMBL; 249769; G840782; -  
 DR EMBL; L20431; G398029; -  
 DR MIM; 143100; -  
 KW DISEASE MUTATION; POLYMORPHISM; TRIPLET REPEAT EXPANSION; APOPTOSIS.  
 FT DOMAIN 18 40 POLY-GLN.  
 FT DOMAIN 41 51 POLY-PRO.  
 FT DOMAIN 65 80 POLY-PRO.  
 FT DOMAIN 1439 1442 POLY-THR.  
 FT DOMAIN 2343 2347 POLY-GLU.  
 FT DOMAIN 2640 2645 POLY-GLU.  
 FT VARIANT 38 40 MISSING.  
 FT SITE 513 514 CLEAVAGE BY APOPAIN (POTENTIAL).  
 FT SITE 530 531 CLEAVAGE BY APOPAIN (POTENTIAL).  
 FT SITE 552 553 CLEAVAGE BY APOPAIN (POTENTIAL).  
 FT SITE 589 590 CLEAVAGE BY APOPAIN (POTENTIAL).  
 FT CONFLICT 2788 2788 V -> I (IN REF. 10).  
 SQ SEQUENCE 3144 AA; 347855 MW; 8BF8AE44 CRC32;

Query Match 68.2%; Score 45; DB 1; Length 3144;

Best Local Similarity 60.0%; Pred. No. 1.27e+01;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 EEEVEVDPT 2352

QY 1 EEOQEVPPDT 10

RESULT 13  
 ID ANKB\_HUMAN STANDARD; PRT; 3924 AA.

AC Q01484; 01-APR-1993 (REL. 25, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)

DE ANKYRIN, BRAIN VARIANT 1 (ANKYRIN B) (ANKYRIN, NONERYTHROID).

GN ANK2.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

[1]

RP SEQUENCE FROM N.A.

RC TISSUE-BRAIN STEM;

RX MEDLINE; 94075409.

RA CHAN W., KORDELI E., BENNETT V.;

RL J. CELL BIOL. 123:1463-1473(1993).

[2]

RP SEQUENCE OF 1-2077 FROM N.A.

RC TISSUE-BRAIN STEM;

RX MEDLINE; 91302466.

RA OTTO E., KUNIMOTO M., MCLAUGHLIN T., BENNETT V.;

RL J. CELL BIOL. 114:241-253(1991).

CC SKELETAL ELEMENTS; THEY BIND TO THE ERYTHROCYTE MEMBRANE PROTEIN

AND TO THE CYTOSKELETAL PROTEINS FODRIN, TUBULIN, VIMENTIN AND

DESMIN. ERYTHROCYTE ANKYRINS ALSO LINK SPECTRIN (BETA CHAIN) TO

THE CYTOPLASMIC DOMAIN OF THE ERYTHROCYTES ANION EXCHANGE PROTEIN;

THEY RETAIN MOST OR ALL OF THESE BINDING FUNCTIONS.

CC TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL

CELLS THROUGHOUT THE BRAIN.

CC PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES

AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE

AND FUNCTION (POTENTIAL).

CC ALTERNATIVE PRODUCTS: THE TWO BRAIN VARIANTS ARE PRODUCED BY

ALTERNATIVE SPLICING OF GENE ANK2.

CC SIMILARITY: CONTAINS 24 ANK REPEATS.

DR EMBL; 226634; G406288; -

DR EMBL; X56957; G29489; -

DR PIR; S14533; S14533.

DR PIR; A39643; A39643.

DR MIM; 106410; -

DR PROSITE; PSS0017; DEATH\_DOMAIN; 1.

KW CYTOSKELETON; ALTERNATIVE SPLICING; REPEAT; ANK REPEAT;  
 KW PHOSPHORYLATION; MULTIGENE FAMILY.  
 FT DOMAIN 63 792 24 X ANK MOTIF REPEATS.  
 FT REPEAT 63 95 ANK MOTIF 1.  
 FT REPEAT 96 128 ANK MOTIF 2.  
 FT REPEAT 129 161 ANK MOTIF 3.  
 FT REPEAT 162 190 ANK MOTIF 4.  
 FT REPEAT 191 231 ANK MOTIF 5.  
 FT REPEAT 232 264 ANK MOTIF 6.  
 FT REPEAT 265 297 ANK MOTIF 7.  
 FT REPEAT 298 330 ANK MOTIF 8.  
 FT REPEAT 331 363 ANK MOTIF 9.  
 FT REPEAT 364 396 ANK MOTIF 10.  
 FT REPEAT 397 429 ANK MOTIF 11.  
 FT REPEAT 430 462 ANK MOTIF 12.  
 FT REPEAT 463 495 ANK MOTIF 13.  
 FT REPEAT 496 528 ANK MOTIF 14.  
 FT REPEAT 529 561 ANK MOTIF 15.  
 FT REPEAT 562 594 ANK MOTIF 16.  
 FT REPEAT 595 627 ANK MOTIF 17.  
 FT REPEAT 628 660 ANK MOTIF 18.  
 FT REPEAT 661 693 ANK MOTIF 19.  
 FT REPEAT 694 726 ANK MOTIF 20.  
 FT REPEAT 727 759 ANK MOTIF 21.  
 FT REPEAT 760 792 ANK MOTIF 22.  
 FT DOMAIN 1773 1950 REPEAT-RICH REGION.  
 FT REPEAT 1773 1784 REPEAT A.  
 FT REPEAT 1785 1796 REPEAT A.  
 FT REPEAT 1797 1808 REPEAT A.  
 FT REPEAT 1809 1820 REPEAT A.  
 FT REPEAT 1821 1832 REPEAT A.  
 FT REPEAT 1833 1844 REPEAT A.  
 FT REPEAT 1845 1856 REPEAT A.  
 FT REPEAT 1857 1867 REPEAT A (APPROXIMATE).  
 FT REPEAT 1868 1879 REPEAT A.  
 FT REPEAT 1880 1891 REPEAT A.  
 FT REPEAT 1892 1902 REPEAT A (APPROXIMATE).  
 FT REPEAT 1903 1914 REPEAT A.  
 FT REPEAT 1915 1926 REPEAT A.  
 FT REPEAT 1927 1938 REPEAT A.  
 FT REPEAT 1939 1950 REPEAT A.  
 FT DOMAIN 3536 3620 DEATH DOMAIN.  
 SQ SEQUENCE 3924 AA; 430337 MW; 0CCC249D CRC32;

Query Match 68.2%; Score 45; DB 1; Length 3924;

Best Local Similarity 40.0%; Pred. No. 1.27e+01;

Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 3839 DMPEIPPET 3848

QY 1 EEOQEVPPDT 10

RESULT 14  
 ID YKEL\_YEAST STANDARD; PRT; 224 AA.  
 AC P36095;  
 DT 01-JUN-1994 (REL. 29, CREATED)  
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 26.2 KD PROTEIN IN SPC42-PTM1 INTERGENIC REGION.  
 GN YKL041W OR YKL254.  
 OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE; 94205268.  
 RA PURNELLE B., TETTELIN H., VAN DYCK L., SKALA J., GOFFEAU A.;  
 RL YEAST 9:1379-1384(1993).  
 DR EMBL; X71621; G666099; -  
 DR EMBL; Z28041; G486052; -  
 DR PIR; S37862; S37862.  
 KW HYPOTHETICAL PROTEIN.

```

SQ SEQUENCE 224 AA; 26242 MW; 3FEAA543 CRC32;
Query Match 66.7%; Score 44; DB 1; Length 224;
Best Local Similarity 55.6%; Pred. No. 2.04e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 194 EEEQEIPDE 202
||:|:|:|
QY 1 EEEQEVPPD 9

RESULT 15
ID DCUP_ECOLI STANDARD; PRT; 354 AA.
AC P29680; P78135;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE UROPORPHYRINOGEN DECARBOXYLASE (EC 4.1.1.37) (UPD).
GN HEME.
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 94040783.
RA NISHIMURA K., INOKUCHI H.;
RL GENE 133:109-113(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 94089392.
RA BLATTNER F.R., BURLAND V.D., PLUNKETT G. III, SOFIA H.J.,
RA DANIELS D.L.;
RL NUCLEIC ACIDS RES. 21:5408-5417(1993).
CC -!- CATALYTIC ACTIVITY: UROPORPHYRINOGEN III - COPROPORPHYRINOGEN +
CC 4 CO(2).
CC -!- PATHWAY: PORPHYRIN BIOSYNTHESIS.
DR EMBL; D12624; G216564; -
DR EMBL; U00006; G409791; -
DR EMBL; AE000473; G2367337; -.
DR PIR; JS0708; JS0708.
DR ECOGENE; EG11543; HEME.
DR PROSITE; PS00906; UROD_1; 1.
DR PROSITE; PS00907; UROD_2; 1.
KW LYASE; DECARBOXYLASE; PORPHYRIN BIOSYNTHESIS.
FT CONFLICT 72 78 AILFSDI -> RSSFRY (IN REF. 1).
FT CONFLICT 83 83 D -> I (IN REF. 1).
FT CONFLICT 89 103 LYFEAGEGPRFTSPV -> SSILKPEKVRVLPQRI
FT (IN REF. 1)
FT CONFLICT 251 256 GGGQWL -> SATVA (IN REF. 1).
SQ SEQUENCE 354 AA; 39248 MW; A9638BF5 CRC32;

Query Match 66.7%; Score 44; DB 1; Length 354;
Best Local Similarity 57.1%; Pred. No. 2.04e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 330 HQDVPE 336
||:|:|:|
QY 3 QOEVPDP 9

```

Search completed: Thu May 13 15:32:55 1999  
Job time : 8 secs.

\*\*\*\*\*  
M A P S R E F  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu May 13 15:33:14 1999; MasPar time 4.01 Seconds  
Tabular output not generated.  
124.180 Million cell updates/sec

Title: >US-09-040-485-7  
Description: (1-10) from US09040485.pep  
Perfect Score: 66  
Sequence: 1 EEQQEVPDPDT 10

Scoring table: PAM 150  
Gap 15

Searched: 165420 seqs, 49795644 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrmb16  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 21.034; Variance 28.456; scale 0.739

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES  |       |             |              |                               |
|------------|-------|-------------|--------------|-------------------------------|
| Result No. | Score | Query Match | Length DB ID | Description                   |
| 1          | 49    | 74.2        | 221 10       | 065441 HYPOTHETICAL 24.8 KD P |
| 2          | 49    | 74.2        | 1406 4       | 015082 KIAA0377               |
| 3          | 49    | 74.2        | 1685 5       | 024754 MASTERMIND             |
| 4          | 48    | 72.7        | 698 4        | 093021 PUTATIVE TUMOR SUPPRES |
| 5          | 47    | 71.2        | 147 2        | 066475 SINGLE STRANDED DNA-BI |
| 6          | 47    | 71.2        | 216 11       | 008904 BRX PROTEIN (FRAGMENT) |
| 7          | 47    | 71.2        | 1120 5       | 020778 SIMILAR TO TRIPLE HELI |
| 8          | 46    | 69.7        | 68 13        | 091773 HEAT SHOCK PROTEIN (HS |
| 9          | 46    | 69.7        | 284 5        | 024013 GS-LIKE ORF'S PROTEIN. |
| 10         | 46    | 69.7        | 360 10       | 040845 OSB28                  |
| 11         | 46    | 69.7        | 646 10       | 038870 CALMODULIN-DOMAIN PROT |
| 12         | 46    | 69.7        | 649 14       | 071093 TERMINAL PROTEIN.      |
| 13         | 46    | 69.7        | 812 5        | 018717 SIMILAR TO S           |
| 14         | 46    | 69.7        | 1320 14      | 006359 150 KD PROTEIN.        |
| 15         | 46    | 69.7        | 1828 14      | 089249 209 KDA READTHROUGH PR |
| 16         | 46    | 69.7        | 7962 4       | 010465 TITIN, SKELETAL MUSCLE |
| 17         | 45    | 68.2        | 370 5        | 002271 F52F12.2               |
| 18         | 45    | 68.2        | 405 6        | 079260 S-ANTIGEN (FRAGMENT)   |
| 19         | 45    | 68.2        | 411 5        | 026766 FLAGELLAR ANTIGEN (FRA |
| 20         | 45    | 68.2        | 974 5        | 092165 ZK270.2C (FRAGMENT).   |

|    |    |      |      |    |        |                        |          |
|----|----|------|------|----|--------|------------------------|----------|
| 21 | 45 | 68.2 | 4343 | 5  | P91847 | ZK270.2A (FRAGMENT).   | 3.69e+01 |
| 22 | 44 | 66.7 | 382  | 14 | Q83924 | E1B LARGE T PROTEIN.   | 5.66e+01 |
| 23 | 44 | 66.7 | 411  | 3  | Q08412 | ORF YOR042W            | 5.66e+01 |
| 24 | 44 | 66.7 | 451  | 1  | Q06022 | SOLUBLE TRANSDUCER PRO | 5.66e+01 |
| 25 | 44 | 66.7 | 584  | 5  | Q19672 | F2IC3.4.               | 5.66e+01 |
| 26 | 44 | 66.7 | 990  | 5  | Q46086 | 1-EVIDENCE-PREDICTED B | 5.66e+01 |
| 27 | 44 | 66.7 | 1225 | 5  | Q17656 | C05C9.3.               | 5.66e+01 |
| 28 | 44 | 66.7 | 1307 | 5  | P91094 | SIMILARITY TO DROSOPHI | 5.66e+01 |
| 29 | 43 | 65.2 | 169  | 5  | Q20621 | F49E2.2.               | 8.62e+01 |
| 30 | 43 | 65.2 | 229  | 2  | P96494 | PUTATIVE GLYCINE DEHYD | 8.62e+01 |
| 31 | 43 | 65.2 | 345  | 14 | Q66098 | P37K PROTEIN.          | 8.62e+01 |
| 32 | 43 | 65.2 | 442  | 4  | Q15579 | TRANSFORMING GROWTH FA | 8.62e+01 |
| 33 | 43 | 65.2 | 664  | 4  | Q92541 | MYELOBLAST KIAA0252 (F | 8.62e+01 |
| 34 | 43 | 65.2 | 664  | 11 | Q70588 | MCASK-A.               | 8.62e+01 |
| 35 | 43 | 65.2 | 725  | 5  | Q26717 | ENTRO 1125 ADG1 AND AD | 8.62e+01 |
| 36 | 43 | 65.2 | 947  | 5  | Q19106 | HYPOTHETICAL PROTEIN F | 8.62e+01 |
| 37 | 43 | 65.2 | 1010 | 5  | Q20818 | F5C5.4.                | 8.62e+01 |
| 38 | 43 | 65.2 | 1012 | 11 | Q64028 | EARLY DEVELOPMENT REGU | 8.62e+01 |
| 39 | 43 | 65.2 | 1549 | 11 | Q60444 | TYPE VII COLLAGEN (FRA | 8.62e+01 |
| 40 | 43 | 65.2 | 1619 | 14 | Q98645 | ORFS 1,2 & 3, STRAIN P | 8.62e+01 |
| 41 | 43 | 65.2 | 1735 | 11 | Q62768 | MUNC13-1.              | 8.62e+01 |
| 42 | 43 | 65.2 | 2150 | 5  | Q17596 | C2SD7.3.               | 8.62e+01 |
| 43 | 43 | 65.2 | 2559 | 5  | Q44113 | PUTATIVE GUANINE NUCLE | 8.62e+01 |
| 44 | 43 | 65.2 | 2912 | 4  | Q14054 | COLLAGEN TYPE VII PREC | 8.62e+01 |
| 45 | 43 | 65.2 | 3473 | 14 | Q83034 | POLYPROTEIN.           | 8.62e+01 |

ALIGNMENTS

| RESULT   | 1  |   |      |      |     |  |  |
|--|--|---|------|------|-----|--|--|
| ID   | 065441   | PRELIMINARY;                            | PRT; | 221  | AA. |  |  |
| AC   | 065441   |   |      |      |     |  |  |
| DT   | 01-AUG-1998  | (TREMBLREL. 07, CREATED)                |      |      |     |  |  |
| DT   | 01-AUG-1998  | (TREMBLREL. 07, LAST SEQUENCE UPDATE)   |      |      |     |  |  |
| DT   | 01-AUG-1998  | (TREMBLREL. 07, LAST ANNOTATION UPDATE) |      |      |     |  |  |
| DE   | HYPOTHETICAL 24.8 KD PROTEIN.                                      |   |      |      |     |  |  |
| GN   | FIC12.195.   |   |      |      |     |  |  |
| OS   | ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).                            |   |      |      |     |  |  |
| OC   | EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;      |   |      |      |     |  |  |
| OC   | CAPPARALES; CRUCIFERAE.  |   |      |      |     |  |  |
| RN   | [1]  |   |      |      |     |  |  |
| RP   | SEQUENCE FROM N.A.   |   |      |      |     |  |  |
| RA   | BEVAN M., TERRY N., ARDILES W., BUYSSHAERT C., DASSEVILLE R.,      |   |      |      |     |  |  |
| RA   | DE CLERCK R., DE KEYSER A., NEYT P., ROUZE P., VAN DEN DAEL H.,    |   |      |      |     |  |  |
| RA   | VILLAROEEL R., GIELEN J., VAN MONTAGU M., BANCROFT I., MEMES H.W., |   |      |      |     |  |  |
| RA   | MAYER K., SCHUELLER C.;  |   |      |      |     |  |  |
| RL   | SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.               |   |      |      |     |  |  |
| RN   | [2]  |   |      |      |     |  |  |
| RP   | SEQUENCE FROM N.A.   |   |      |      |     |  |  |
| RA   | EU ARABIDOPSIS SEQUENCING PROJECT;                                 |   |      |      |     |  |  |
| RL   | SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.               |   |      |      |     |  |  |
| DR   | EMBL; AL02224; E1283977; -.  |   |      |      |     |  |  |
| KW   | HYPOTHETICAL PROTEIN.  |   |      |      |     |  |  |
| SQ   | SEQUENCE 221 AA; 24834 MW; 7BEFAC97 CRC32;                         |   |      |      |     |  |  |
| Query Match 74.2%; Score 49; DB 10; Length 221;            |  |   |      |      |     |  |  |
| Best Local Similarity 60.0%; Pred. No. 6.28e+00;           |  |   |      |      |     |  |  |
| Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0; |  |   |      |      |     |  |  |
| Db   | 14   | EEQESPSPES                              | 23   |      |     |  |  |
|  |  |   |      |      |     |  |  |
| Qy   | 1  | EEQQEVPDPDT                             | 10   |      |     |  |  |
| RESULT 2   |  |   |      |      |     |  |  |
| ID   | 015082   | PRELIMINARY;                            | PRT; | 1406 | AA. |  |  |
| AC   | 015082   |   |      |      |     |  |  |
| DT   | 01-JAN-1998  | (TREMBLREL. 05, CREATED)                |      |      |     |  |  |
| DT   | 01-JAN-1998  | (TREMBLREL. 05, LAST SEQUENCE UPDATE)   |      |      |     |  |  |
| DT   | 01-JUN-1998  | (TREMBLREL. 06, LAST ANNOTATION UPDATE) |      |      |     |  |  |
| DE   | KIAA0377.  |   |      |      |     |  |  |
| GN   | KIAA0377.  |   |      |      |     |  |  |
| OS   | HOMO SAPIENS (HUMAN).  |   |      |      |     |  |  |

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE; 97349984.  
 RA NAGASE T., ISHII K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N.,  
 RA TANAKA A., KOTANI H., NOMURA N., OHARA O.,  
 RL DNA RES. 4:141-150(1997).  
 DR EMBL; AB002375; D1021673;  
 SQ SEQUENCE 1406 AA; 156319 MW; 68F9BC7E CRC32;

Query Match 74.2%; Score 49; DB 4; Length 1406;  
 Best Local Similarity 55.6%; Pred. No. 6.28e+00;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 44 DEDEVPPPE 52  
 :|:|:|:|:  
 Qy 1 EEQEVPPD 9

RESULT 3  
 ID Q24754 PRELIMINARY; PRT; 1655 AA.  
 AC Q24754;  
 DT 01-NOV-1996 (TREMREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMREL. 01, LAST ANNOTATION UPDATE)  
 DE MASTERMIND.  
 OS DROSOPHILA VIRILIS. (FRUIT FLY).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
 RN [1]  
 RP SEQUENCE OF 260-762 FROM N.A.  
 RX MEDLINE; 91251140.  
 RA NEWFELD S.J., SMOLLER D.A., YEDVOBNICK B.;  
 RL J. MOL. EVOL. 32:415-420(1991).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94111143.  
 RA NEWFELD S.J., SCHMID A.T., YEDVOBNICK B.;  
 RL J. MOL. EVOL. 37:483-495(1993).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94365848.  
 RA NEWFELD S.J., TACHIDA H., YEDVOBNICK B.;  
 RL J. MOL. EVOL. 38:637-641(1994).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RA NEWFELD S.J.;  
 RL SUBMITTED (MAY-1992) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; M92914; G157834; -;  
 DR FLYBASE; FBgn0013119; Dvir\man.  
 SQ SEQUENCE 1655 AA; 175048 MW; D521E17E CRC32;

Query Match 74.2%; Score 49; DB 5; Length 1655;  
 Best Local Similarity 55.6%; Pred. No. 6.28e+00;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 1084 QQQQVPPN 1092  
 :|:|:|:|:  
 Qy 1 EEQEVPPD 9

RESULT 4  
 ID Q93021 PRELIMINARY; PRT; 698 AA.  
 AC Q93021;  
 DT 01-FEB-1997 (TREMREL. 02, CREATED)  
 DT 01-FEB-1997 (TREMREL. 02, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)  
 DE PUTATIVE TUMOR SUPPRESSOR.  
 GN LUCAL5.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-BRAIN;  
 RA SIMMLER M.C., HEARD E., ROUGEULLE C., CRUAUD C., WEISSENBACH J.,

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BENILEY D., MAGGI L.;  
 RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA WATERSTON R.;  
 RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; U73168; G1613900; -;  
 DR PFAM; PF00641; zf-RanBP.  
 SQ SEQUENCE 698 AA; 78066 MW; 36A3242B CRC32;

Query Match 72.7%; Score 48; DB 4; Length 698;  
 Best Local Similarity 60.0%; Pred. No. 9.87e+00;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 94 DSEQEVPPGT 103  
 :|:|:|:|:  
 Qy 1 EEQEVPPDT 10

RESULT 5  
 ID O66475 PRELIMINARY; PRT; 147 AA.  
 AC O66475;  
 DT 01-AUG-1998 (TREMREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
 DE SINGLE STRANDED DNA-BINDING PROTEIN.  
 GN SSB.  
 OS AQUIFEX AEOLICUS.  
 OC EUBACTERIA; AQUIFICALES; AQUIFEXACEAE; AQUIFEX.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VF5;  
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
 RL NATURE 392:353-358(1998).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VF5;  
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; AE000672; G2982816; -;  
 KW DNA-BINDING.  
 SQ SEQUENCE 147 AA; 17132 MW; BFA26F97 CRC32;

Query Match 71.2%; Score 47; DB 2; Length 147;  
 Best Local Similarity 75.0%; Pred. No. 1.54e+01;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 115 EEEEVPPP 122  
 :|:|:|:|:  
 Qy 1 EEQEVPPP 8

RESULT 6  
 ID O08904 PRELIMINARY; PRT; 216 AA.  
 AC O08904;  
 DT 01-JUL-1997 (TREMREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
 DE BRX PROTEIN (FRAGMENT).  
 GN BRX.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-BRAIN;  
 RA SIMMLER M.C., HEARD E., ROUGEULLE C., CRUAUD C., WEISSENBACH J.,

RA AVNER P.;  
RL MAMM. GENOME 0:0-0(0).  
DR EMBL; Y11896; E311743; -  
FT NON\_TER 1  
SQ SEQUENCE 216 AA; 24385 MW; 47AD381A CRC32;  
Query Match 71.2%; Score 47; DB 11; Length 216;  
Best Local Similarity 75.0%; Pred. No. 1.54e+01;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 76 EEEEEVPP 83  
QY 1 EEQEVPP 8  
RESULT 7  
ID Q20778 PRELIMINARY; PRT; 1120 AA.  
AC Q20778;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE SIMILAR TO TRIPLE HELICAL REGION OF COLLAGENS.  
GN F54B8.1.  
OS CAENORHABDITIS ELGANS.  
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA BURTON J., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
RA WILSON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,  
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,  
RA VAUDIN M., VAUGHAN K., SULSTON J., THIERRY-MIEG J., THOMAS K.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.;  
RL NATURE 0:0-0(0).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; U12966; G529221; -  
SQ SEQUENCE 1120 AA; 119368 MW; E0352B8E CRC32;  
Query Match 71.2%; Score 47; DB 5; Length 1120;  
Best Local Similarity 50.0%; Pred. No. 1.54e+01;  
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
Db 127 EEEEEVPAEA 136  
QY 1 EEQEVPPDT 10  
RESULT 8  
ID Q91773 PRELIMINARY; PRT; 68 AA.  
AC Q91773;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE HEAT SHOCK PROTEIN (HSP30) (FRAGMENT).  
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 84221917.

RA BIENZ M.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 81:3138-3142(1984).  
DR EMBL; K02305; G214267; -  
DR PFAM; PF00011; HSP20.  
KW HEAT SHOCK.  
FT NON\_TER 1  
SQ SEQUENCE 68 AA; 7405 MW; 4D1E326C CRC32;  
Query Match 69.7%; Score 46; DB 13; Length 68;  
Best Local Similarity 55.6%; Pred. No. 2.39e+01;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 48 DAQEIPDDA 56  
QY 2 EQQEVPPDT 10  
RESULT 9  
ID Q24013 PRELIMINARY; PRT; 264 AA.  
AC Q24013;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE G5-LIKE ORF'S PROTEIN.  
OS DICTYOSTELIUM MUCOROIDES.  
OC PLASMID DMP1; NUCLEAR.  
OC EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DMUC2;  
RX MEDLINE; 94302132.  
RA KIYOSAWA H., HUGHES J.E., WELKER D.L.;  
RL PLASMID 31:121-130(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DMUC2;  
RX MEDLINE; 98198836.  
RA RIEBEN W.K. JR., GONZALES C.M., GONZALES S.T., PILKINGTON K.J.,  
RA KIYOSAWA H., HUGHES J.E., WELKER D.L.;  
RL GENETICS 148:1117-1125(1998).  
DR EMBL; U00175; G2246436; -  
KW PLASMID.  
SQ SEQUENCE 264 AA; 30603 MW; 25AD28AD CRC32;  
Query Match 69.7%; Score 46; DB 5; Length 264;  
Best Local Similarity 62.5%; Pred. No. 2.39e+01;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
Db 197 DEEEVPP 204  
QY 1 EQQEVPP 8  
RESULT 10  
ID Q40645 PRELIMINARY; PRT; 360 AA.  
AC Q40645;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE OSB28.  
OS ORYZA SATIVA (RICE).  
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;  
OC CYPERALES; GRAMINEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NIPPONBARE;  
RX MEDLINE; 96417817.  
RA NAKAGAWA H., OHMIYA K., HATTORI T.;  
RL PLANT J. 9:217-227(1996).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.  
DR EMBL; U42208; G1147632; -



2  
1  
1  
2

RP SEQUENCE FROM N.A.  
RC STRAIN-US-NEBRASKA, 1981 WILD-TYPE;  
RX MEDLINE: 93303914.  
RA SHIRAKO Y., WILSON M.A.;  
RL VIROLOGY 195:16-32(1993).  
DR EMBL: L07937; G295509; .  
SQ SEQUENCE 1320 AA; 149976 MW; 58565281 CRC32;

Query Match 69.7%; Score 46; DB 14; Length 1320;  
Best Local Similarity 44.4%; Pred. No. 2.39e+01;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Db 161 DEKEEIPPE 169  
:|:|:|:  
QY 1 EEQVEVPPD 9

RESULT 15  
ID Q89249 PRELIMINARY; PRT; 1828 AA.  
AC Q89249;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE 209 KDA READTHROUGH PROTEIN.  
OS SOIL-BORNE WHEAT MOSAIC VIRUS.  
OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; ROD-SHAPED SS-RNA VIRUSES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-US-NEBRASKA, 1981 WILD-TYPE;  
RX MEDLINE: 93303914.  
RA SHIRAKO Y., WILSON M.A.;  
RL VIROLOGY 195:16-32(1993).  
DR EMBL: L07937; G295508; .  
SQ SEQUENCE 1828 AA; 209203 MW; ABA060DD CRC32;

Query Match 69.7%; Score 46; DB 14; Length 1828;  
Best Local Similarity 44.4%; Pred. No. 2.39e+01;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Db 161 DEKEEIPPE 169  
:|:|:|:  
QY 1 EEQVEVPPD 9

Search completed: Thu May 13 15:33:44 1999  
Job time : 30 secs.

\*\*\*\*\*  
M P S R C H  
\*\*\*\*\*  
(TIM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu May 13 15:35:46 1999; Maspar time 2.54 Seconds  
Tabular output not generated.  
63.693 Million cell updates/sec

Title: >US-09-040-485-8  
Description: (1-10) from US09040485.pep (1 of 2)  
Perfect Score: 67  
Sequence: 1 DGPTGEPQQE 10

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq32  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 14.719; Variance 44.118; scale 0.334

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | ID      | Description           | Pred. No. |
|------------|-------|-------------|---------|-----------------------|-----------|
| 1          | 47    | 70.1        | W14053  | TIM01 mutant protein. | 5.92e+01  |
| 2          | 47    | 70.1        | W14052  | TIM protein.          | 5.92e+01  |
| 3          | 47    | 70.1        | W14051  | TIM protein splice va | 5.92e+01  |
| 4          | 47    | 70.1        | R44431  | eryA region polypepti | 5.92e+01  |
| 5          | 45    | 67.2        | 1341 13 | Collagen alpha 1 (I)  | 9.86e+01  |
| 6          | 45    | 67.2        | 1418 13 | Collagen alpha 1 (II) | 9.86e+01  |
| 7          | 45    | 67.2        | R59751  | Type II collagen.     | 9.86e+01  |
| 8          | 44    | 65.7        | W2310   | ELA 243 amino acid pr | 1.27e+02  |
| 9          | 44    | 65.7        | W03560  | Human adrenoleucodyst | 1.27e+02  |
| 10         | 44    | 65.7        | W46110  | Human ALD             | 1.27e+02  |
| 11         | 43    | 64.2        | W43039  | A NADPH oxidase deriv | 1.63e+02  |
| 12         | 43    | 64.2        | P93285  | Sequence of clone HIV | 1.63e+02  |
| 13         | 43    | 64.2        | 3210 1  | Deduced sequence enco | 1.63e+02  |
| 14         | 42    | 62.7        | R06684  | Human papilloma virus | 2.09e+02  |
| 15         | 42    | 62.7        | R14295  | Seroreactive epitope  | 2.09e+02  |
| 16         | 42    | 62.7        | R25268  | HPV vaccine antigenic | 2.09e+02  |
| 17         | 42    | 62.7        | R31213  | HPV-16 E7 peptide.    | 2.09e+02  |
| 18         | 42    | 62.7        | W46886  | Amino acid sequence o | 2.09e+02  |

|    |    |      |         |        |                       |          |
|----|----|------|---------|--------|-----------------------|----------|
| 19 | 42 | 62.7 | 98 8    | R42361 | Human papillomavirus  | 2.09e+02 |
| 20 | 42 | 62.7 | 98 4    | R22767 | HPV E7 peptide.       | 2.09e+02 |
| 21 | 42 | 62.7 | 172 20  | R97563 | Human papilloma virus | 2.09e+02 |
| 22 | 42 | 62.7 | 172 20  | R97562 | Human papilloma virus | 2.09e+02 |
| 23 | 42 | 62.7 | 181 17  | R34589 | Human R33 platelet de | 2.09e+02 |
| 24 | 42 | 62.7 | 181 17  | R94588 | Rat R33 platelet deri | 2.09e+02 |
| 25 | 42 | 62.7 | 200 15  | R76871 | Talin C-terminal pept | 2.09e+02 |
| 26 | 42 | 62.7 | 262 5   | R27724 | HPV 16 E7 protein fra | 2.09e+02 |
| 27 | 42 | 62.7 | 266 20  | R97561 | Human papilloma virus | 2.09e+02 |
| 28 | 42 | 62.7 | 324 24  | W24971 | Rat amelin-2 protein. | 2.09e+02 |
| 29 | 42 | 62.7 | 407 24  | W24970 | Rat amelin-1 protein. | 2.09e+02 |
| 30 | 42 | 62.7 | 433 25  | W05389 | Mouse SH3P7 protein.  | 2.09e+02 |
| 31 | 42 | 62.7 | 1442 14 | R79480 | Rat type II collagen. | 2.09e+02 |
| 32 | 42 | 62.7 | 1931 5  | R27649 | Human calcium channel | 2.09e+02 |
| 33 | 42 | 62.7 | 2237 14 | R71006 | Human neuronal calciu | 2.09e+02 |
| 34 | 42 | 62.7 | 2237 6  | R33550 | Sequence of the alpha | 2.09e+02 |
| 35 | 42 | 62.7 | 2339 14 | R71005 | Human neuronal calciu | 2.09e+02 |
| 36 | 42 | 62.7 | 2339 6  | R33549 | Sequence of the alpha | 2.09e+02 |
| 37 | 41 | 61.2 | 92 24   | W14586 | Streptococcus pneumon | 2.68e+02 |
| 38 | 41 | 61.2 | 623 24  | W12843 | Pro-alpha1(III):(I) C | 2.68e+02 |
| 39 | 41 | 61.2 | 626 24  | W12842 | Truncated pro-alpha1( | 2.68e+02 |
| 40 | 41 | 61.2 | 806 5   | R27481 | RP-III residual prote | 2.68e+02 |
| 41 | 41 | 61.2 | 1070 14 | R77083 | M.vulgaris alpha-1,4- | 2.68e+02 |
| 42 | 41 | 61.2 | 1070 13 | R72712 | Alpha-1,4-glucan lyas | 2.68e+02 |
| 43 | 41 | 61.2 | 1078 13 | R71704 | Collagen alpha 1 (III | 2.68e+02 |
| 44 | 41 | 61.2 | 1838 10 | R53257 | Human collagen (Type  | 2.68e+02 |
| 45 | 41 | 61.2 | 2353 19 | R99393 | Haemophilus adhesin   | 2.68e+02 |

ALIGNMENTS

RESULT 1  
ID W14053 standard; Protein; 749 AA.

AC W14053;  
DT 28-MAY-1997 (first entry)

DE TIM01 mutant protein.

KW TIM; timeless protein; nuclear translocation protein; circadian rhythm;  
KW sleep-wake cycle; light sensitive; environmental cycle; nuclear protein;  
KW Drosophila period gene; transcription factor; PAS domain; depression;  
KW narcolepsy; PER; jet lag; NTP; therapy.  
OS Drosophila melanogaster.

FS Key Location/Qualifiers  
FT region 351..380  
FT /note="acidic region"

FT region 540..553  
FT /note="basic region, possible nuclear localisation  
signal"

WO9629406-A2.  
26-SEP-1996. U03830.  
20-MAR-1996; US-408518.  
16-MAY-1995; US-442214.  
02-NOV-1995; US-552354.  
(UYPE-) UNIV PENNSYLVANIA.  
(UYRQ) UNIV ROCKEFELLER.

PI Myers MP, Price JL, Sehgal A, Vossall LB, Young MW;  
DR WPI; 96-443182/44.  
DR N-PSDB; T60336.

PT Nuclear translocation protein which binds to protein involved in  
PT circadian rhythms - used to develop prods. for diagnosis, prevention  
PT or treatment of disorders associated with circadian rhythms, e.g.

PT jetlag and narcolepsy  
PS Example 9; : 131pp; English.

CC This sequence represents the 01 mutant of the Drosophila melanogaster  
CC "timeless" (TIM) protein. The TIM01 mutant is an arrhythmic mutation. TIM  
CC is a nuclear translocation protein (NTP) of the invention. The NTPs of  
CC the invention have specific binding activity to a protein involved in  
CC circadian rhythms, and cyclic transcription patterns related to the  
CC sleep-wake cycle. The NTP is also preferably light sensitive, and has a  
CC stabilising effect on the circadian rhythm protein. The NTP also has the  
CC ability to aid the process of circadian rhythm entrainment to  
CC environmental cycles of light. TIM has specific binding activity for the  
CC Drosophila period (PER) gene. PER is a nuclear protein which has homology

CC to the family of transcription factors containing the PAS domain, and is  
 CC a protein involved in circadian rhythms, but the biological function of  
 CC PER is unknown. The amount of PER fluctuates with a circadian rhythm.  
 CC and the protein is also phosphorylated with a circadian rhythm. The NTPs  
 CC of the invention, their fragments, agonists, mimics and antagonists are  
 CC useful for preventing and/or treating disorders of a circadian rhythm  
 CC such as depression, narcolepsy or jet lag. The products can also be used  
 CC for detecting and/or measuring conditions so as to classify groups of  
 CC individuals with the disorders.  
 SQ Sequence 749 AA;

Query Match 70.1%; Score 47; DB 21; Length 749;  
 Best Local Similarity 60.0%; Pred. No. 5.92e+01;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 510 dpggkpkqh 519  
 ||| | | | |  
 QY 1 DGTGEPOQE 10

RESULT 2  
 ID W14052 standard; Protein; 1122 AA.

AC W14052;  
 DT 28-MAY-1997 (first entry)  
 DE TIM protein.  
 KW TIM; timeless protein; nuclear translocation protein; circadian rhythm;  
 KW sleep-wake cycle; light sensitive; environmental cycle; nuclear protein;  
 KW Drosophila period gene; transcription factor; PAS domain; depression;  
 KW narcolepsy; PER; jet lag; NTP; therapy.  
 OS Drosophila melanogaster.  
 FH Key Location/Qualifiers  
 FT region 351..380  
 FT /note="acidic region"  
 FT region 540..553  
 FT /note="basic region, possible nuclear localisation  
 FT signal"

WO9629406-A2.  
 26-SEP-1996.  
 PD 20-MAR-1996; US03830.  
 PF 20-MAR-1995; US-408518.  
 PR 16-MAY-1995; US-442214.  
 PR 02-NOV-1995; US-552354.  
 PR (UYPE-) UNIV PENNSYLVANIA.  
 PA (UYRQ) UNIV ROCKEFELLER.  
 PI Myers MP, Price JL, Sehgal A, Voshall LB, Young MW;  
 DR N-PSDB; T60333, T60335.  
 DR WPI; 96-443182/44.  
 PT Nuclear translocation protein which binds to protein involved in  
 PT circadian rhythms - used to develop prods. for diagnosis, prevention  
 PT or treatment of disorders associated with circadian rhythms, e.g.  
 PT jetlag and narcolepsy  
 PS Claim 3; Fig 11; 131pp; English.  
 CC This sequence represents the Drosophila melanogaster "timeless" (TIM)  
 CC protein. TIM is a nuclear translocation protein (NTP) of the invention.  
 CC The NTPs of the invention have specific binding activity to a protein  
 CC involved in circadian rhythms, and cyclic transcription patterns related  
 CC to the sleep-wake cycle. The NTP is also preferably light sensitive, and  
 CC has a stabilising effect on the circadian rhythm protein. The NTP also  
 CC has the ability to aid the process of circadian rhythm entrainment to  
 CC environmental cycles of light. TIM has specific binding activity for the  
 CC Drosophila period (PER) gene. PER is a nuclear protein which has homology  
 CC to the family of transcription factors containing the PAS domain, and is  
 CC a protein involved in circadian rhythms, but the biological function of  
 CC PER is unknown. The amount of PER fluctuates with a circadian rhythm.  
 CC and the protein is also phosphorylated with a circadian rhythm. The NTPs  
 CC of the invention, their fragments, agonists, mimics and antagonists are  
 CC useful for preventing and/or treating disorders of a circadian rhythm  
 CC such as depression, narcolepsy or jet lag. The products can also be used  
 CC for detecting and/or measuring conditions so as to classify groups of  
 CC individuals with the disorders.  
 SQ Sequence 1122 AA;

Query Match 70.1%; Score 47; DB 21; Length 1122;

Best Local Similarity 60.0%; Pred. No. 5.92e+01;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Db 510 dpggkpkqh 519  
 ||| | | | |  
 QY 1 DGTGEPOQE 10

RESULT 3  
 ID W14051 standard; Protein; 1389 AA.

AC W14051;  
 DT 28-MAY-1997 (first entry)  
 DE TIM protein splice variant.  
 KW TIM; timeless protein; nuclear translocation protein; circadian rhythm;  
 KW sleep-wake cycle; light sensitive; environmental cycle; nuclear protein;  
 KW Drosophila period gene; transcription factor; PAS domain; depression;  
 KW narcolepsy; PER; jet lag; NTP; therapy.  
 OS Drosophila melanogaster.  
 FH Key Location/Qualifiers  
 FT region 351..380  
 FT /note="acidic region"  
 FT region 540..553  
 FT /note="basic region, possible nuclear localisation  
 FT signal"

WO9629406-A2.  
 26-SEP-1996.  
 PD 20-MAR-1996; US03830.  
 PF 20-MAR-1995; US-408518.  
 PR 16-MAY-1995; US-442214.  
 PR 02-NOV-1995; US-552354.  
 PR (UYPE-) UNIV PENNSYLVANIA.  
 PA (UYRQ) UNIV ROCKEFELLER.  
 PI Myers MP, Price JL, Sehgal A, Voshall LB, Young MW;  
 DR N-PSDB; T60334.  
 DR WPI; 96-443182/44.  
 PT Nuclear translocation protein which binds to protein involved in  
 PT circadian rhythms - used to develop prods. for diagnosis, prevention  
 PT or treatment of disorders associated with circadian rhythms, e.g.  
 PT jetlag and narcolepsy  
 PS Claim 3; Fig 11; 131pp; English.  
 CC This sequence represents the splice variant of the Drosophila  
 CC melanogaster "timeless" (TIM) protein. TIM is a nuclear translocation  
 CC protein (NTP) of the invention. The NTPs of the invention have specific  
 CC binding activity to a protein involved in circadian rhythms, and cyclic  
 CC transcription patterns related to the sleep-wake cycle. The NTP is also  
 CC preferably light sensitive, and has a stabilising effect on the circadian  
 CC rhythm protein. The NTP also has the ability to aid the process of  
 CC circadian rhythm entrainment to environmental cycles of light. TIM has  
 CC specific binding activity for the Drosophila period (PER) gene. PER is a  
 CC nuclear protein which has homology to the family of transcription factors  
 CC containing the PAS domain, and is a protein involved in circadian  
 CC rhythms, but the biological function of PER is unknown. The amount of  
 CC PER fluctuates with a circadian rhythm, and the protein is also  
 CC phosphorylated with a circadian rhythm. The NTPs of the invention, their  
 CC fragments, agonists, mimics and antagonists are useful for preventing  
 CC and/or treating disorders of a circadian rhythm such as depression,  
 CC narcolepsy or jet lag. The products can also be used for detecting and/or  
 CC measuring conditions so as to classify groups of individuals with the  
 CC disorders.  
 SQ Sequence 1389 AA;

Query Match 70.1%; Score 47; DB 21; Length 1389;  
 Best Local Similarity 60.0%; Pred. No. 5.92e+01;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 510 dpggkpkqh 519  
 ||| | | | |  
 QY 1 DGTGEPOQE 10

RESULT 4  
 ID R44431 standard; Protein; 3567 AA.  
 AC R44431;

DT 22-DEC-1993 (first entry)  
 DE eryA region polypeptide module #2.  
 KW Saccaropolyspora erythraea; eryA; biosynthesis; polyketide; module;  
 KW erythromycin; condensation; elongation; acyl chain growth;  
 KW gene replacement.  
 OS Saccaropolyspora erythraea.  
 PN W09313663-A.  
 PD 22-JUL-1993.  
 PF 17-JAN-1992; U00427.  
 PR 17-JAN-1992; WO-U00427.  
 PA (ABBO ) ABBOTT LAB.  
 PI Donadio S, Katz L, McAlpine JB;  
 DR WPI: 93-242804/30.  
 DR N-PSDB; Q46806.  
 PT Biosynthesis of specific polyketide analogues esp. erythromycin  
 PT cpds. - by introducing altered biosynthetic gene-contg. DNA into  
 PT microorganisms  
 PS Disclosure: Fig 2; 133pp; English.  
 CC The sequences given in R4430-32 are encoded by the eryA fragment of  
 CC the Saccaropolyspora erythraea genome. These polypeptides are  
 CC involved in the biosynthesis of the polyketide segment of erythromycin.  
 CC eryA is organised in modules and each module takes care of one  
 CC condensation step. The precise succession of elongation steps is  
 CC dictated by the genetic order of the modules. The DNA encoding  
 CC these polypeptides may be specifically altered such that novel  
 CC polyketide molecules of desired structure are produced. Three types  
 CC of alteration may be produced; those inactivating a single function in  
 CC a module which does not arrest acyl chain growth; those inactivating a  
 CC single function in a module which does affect chain growth; and those  
 CC affecting an entire module. The mutations may be introduced by gene  
 CC replacement.  
 SQ Sequence 3567 AA;

Query Match 70.1%; Score 47; DB 8; Length 3567;  
 Best Local Similarity 87.3%; Pred. No. 5.92e+01;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1376 gptgepaq 1383  
 QY 2 GPTGEPOQ 9

RESULT 5  
 ID R71701 standard; protein; 1341 AA.  
 AC R71701;  
 DT 17-OCT-1995 (first entry)  
 DE Collagen alpha 1 (I) chain precursor.  
 KW Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;  
 KW disorder; osteoporosis; metastatic progression; Paget's disease;  
 KW hyperthyroidism; bone; resorption; rheumatoid arthritis;  
 KW osteoarthritis; vasculitis syndrome.  
 OS Homo sapiens.  
 FT Key  
 FT misc\_difference 2028  
 FT Location/Qualifiers  
 FT /note= "Unidentified amino acid."  
 PN W09508115-A.  
 PD 23-MAR-1995.  
 PF 19-SEP-1994; DK0348.  
 PR 17-SEP-1993; DK-001040.  
 PA (OSTE-) OSTEOMETER AS.  
 PI Bonde M, Qvist P;  
 DR WPI: 95-131456/17.  
 PT Assaying collagen fragments in body fluid by immunoassay - using  
 PT antibodies raised against synthetic peptide(s) contg. potential  
 PT crosslinking sites, to diagnose and monitor disorders of collagen  
 PT metabolism, e.g. osteoporosis.  
 PS Disclosure (Appendix A); Page 49; 87pp; English.  
 CC Determination of collagen fragments in body fluids can be achieved  
 CC by immunoassay using antibodies directed against synthetic peptides  
 CC derived from collagen which contain sites of potential crosslinking.  
 CC The method is used to diagnose and monitor treatment of disorders of  
 CC collagen metabolism (degradation of type I collagen may indicate  
 CC osteoporosis, metastatic progression, Paget's disease,

CC hyperthyroidism or other conditions involving excessive bone  
 CC resorption; degradation of type II collagen may indicate rheumatoid  
 CC arthritis or osteoarthritis; and of type III collagen, vaculitis  
 CC syndrome). The method can also be used to assess the toxicity of a  
 CC compound and to test drugs for their effect on collagen metabolism.  
 SQ Sequence 1341 AA;

Query Match 67.2%; Score 45; DB 13; Length 1341;  
 Best Local Similarity 44.4%; Pred. No. 9.86e+01;  
 Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 202 zgpgpgpgz 210  
 QY 1 DGPTGEPOQ 9

RESULT 6  
 ID R71703 standard; protein; 1418 AA.  
 AC R71703;  
 DT 17-OCT-1995 (first entry)  
 DE Collagen alpha 1 (II) chain precursor.  
 KW Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;  
 KW disorder; osteoporosis; metastatic progression; Paget's disease;  
 KW hyperthyroidism; bone; resorption; rheumatoid arthritis;  
 KW osteoarthritis; vasculitis syndrome.  
 OS Homo sapiens.  
 PN W09508115-A.  
 PD 23-MAR-1995.  
 PF 19-SEP-1994; DK0348.  
 PR 17-SEP-1993; DK-001040.  
 PA (OSTE-) OSTEOMETER AS.  
 PI Bonde M, Qvist P;  
 DR WPI: 95-131456/17.  
 PT Assaying collagen fragments in body fluid by immunoassay - using  
 PT antibodies raised against synthetic peptide(s) contg. potential  
 PT crosslinking sites, to diagnose and monitor disorders of collagen  
 PT metabolism, e.g. osteoporosis.  
 PS Disclosure (Appendix A); Page 53; 87pp; English.  
 CC Determination of collagen fragments in body fluids can be achieved  
 CC by immunoassay using antibodies directed against synthetic peptides  
 CC derived from collagen which contain sites of potential crosslinking.  
 CC The method is used to diagnose and monitor treatment of disorders of  
 CC collagen metabolism (degradation of type I collagen may indicate  
 CC osteoporosis, metastatic progression, Paget's disease,  
 CC hyperthyroidism or other conditions involving excessive bone  
 CC resorption; degradation of type II collagen may indicate rheumatoid  
 CC arthritis or osteoarthritis; and of type III collagen, vaculitis  
 CC syndrome). The method can also be used to assess the toxicity of a  
 CC compound and to test drugs for their effect on collagen metabolism.  
 SQ Sequence 1418 AA;

Query Match 67.2%; Score 45; DB 13; Length 1418;  
 Best Local Similarity 66.7%; Pred. No. 9.86e+01;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 930 gpgpgpgq 938  
 QY 2 GPTGEPOQ 10

RESULT 7  
 ID R59751 standard; protein; 1418 AA.  
 AC R59751;  
 DT 14-FEB-1995 (first entry)  
 DE Type II collagen.  
 KW Collagen; triple helix; articular cartilage; collagenase;  
 KW degradation; monoclonal antibody; epitope; matrix;  
 KW metalloproteinase.  
 OS Homo sapiens.  
 FT Key  
 FT Location/Qualifiers  
 FT peptide 1..24  
 FT /label= Signal peptide.  
 PN W09414070-A.

PD 23-JUN-1994.  
 PF 06-DEC-1993; CA0522.  
 PR 04-DEC-1992; US-984123.  
 PA (SHRI-) SHRINERS HOSPITALS FOR CRIPPLED CHILDREN.  
 PI Hollander AB, Poole AR;  
 DR WPI: 94-234222/28.  
 PT Determin. of cartilage degradation - using a monoclonal antibody  
 to measure the amt. of unwound collagen or fragments in samples  
 PS Disclosure: Figure 1; 119pp; English.  
 CC Type II collagen constitutes the bulk of the fibrillar backbone of  
 cartilage matrix. It is composed of a tightly wound triple helix  
 which can only be cleaved by the metalloproteinase collagenase to  
 produce 3/4 and 1/4 length alpha chain fragments. The destruction of  
 articular cartilage is due, in part, to the degradation of collagen.  
 CC Incapable of maintaining its helical structure at physiological  
 temperatures, collagenase-cleaved collagens unwind and become  
 susceptible to further degradation by other proteinases. By  
 CC producing monoclonal antibodies directed against epitopes which are  
 only revealed when collagen is unwound, the antibodies provide a  
 CC means of determining the degradation of cartilage in a biological  
 CC sample. The antibodies do not bind to native helical collagen.  
 CC Epitopes used in the production of such antibodies are described in  
 CC R59749, R59750 and R67742.  
 SQ Sequence 1418 AA;

Query Match 67.2%; Score 45; DB 11; Length 1418;  
 Best Local Similarity 66.7%; Pred. No. 9.86e+01;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 930 gspgpgqq 938  
 QY 2 GPTGEPOQE 10  
 :|||:|

RESULT 8  
 ID R52601 standard; Protein; 243 AA.  
 AC R52601;  
 DT 05-DEC-1994 (first entry)  
 DE EIA 243 amino acid protein.  
 KW Alternative splicing; EIA; adenovirus; early gene; oncogenes;  
 transcription regulation protein; oncogenic transformation; rodent;  
 KW phenotype; human; tumour cell; contact inhibition; differentiation;  
 KW adenovirus 5; anchorage-independent growth; tumorigenic potential;  
 KW reorganisation; flat morphology; tumour suppressor.  
 OS Adenovirus.  
 PN W09409160-A.  
 PD 28-APR-1994.  
 PR 13-OCT-1992; U09774.  
 PA (LJOL-) LA JOLLA CANCER RES FOUND.  
 PI Frisch SW;  
 DR WPI: 94-151344/18.  
 DR N-PSDB; Q62526.  
 PT Restoring contact inhibition to hyperproliferative cells - by  
 introducing nucleic acid encoding EIA peptide, also promoting  
 PT differentiation, used for treating malignancies  
 PS Disclosure: Page 20-22; 41pp; English.  
 CC This sequence is encoded by the product of alternative splicing of the  
 CC EIA RNA such that the information contained within this sequence is a  
 CC subset of the larger EIA protein, see also R52602. EIA is an adeno-  
 CC virus early gene and produces two products. The 243 and 289 residue  
 CC proteins are both transcriptional regulation proteins which facilitate  
 CC the oncogenic transformation of certain rodent cells by other  
 CC oncogenes. The adenovirus EIA gene unexpectedly influences the  
 CC phenotype of human tumour cells so as to restore their contact  
 CC inhibitory properties and promote differentiation. Stable-expression  
 CC of the adenovirus 5 EIA gene reduces anchorage-independent growth and  
 CC tumorigenic potential, promotes reorganisation, induces flat  
 CC morphology, and restores contact inhibition in human tumour cell  
 CC lines. Therefore EIA acts as a tumour suppressor gene in this human  
 CC context.  
 SQ Sequence 243 AA;

Query Match 65.7%; Score 44; DB 10; Length 243;  
 Best Local Similarity 50.0%; Pred. No. 1.27e+02;  
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 138 egpvsepepe 147  
 QY 1 DGPTGEPOQE 10  
 :||:|:

RESULT 9  
 ID W03560 standard; Protein; 745 AA.  
 AC W03560;  
 DT 26-FEB-1997 (first entry)  
 DE Human adrenoleucodystrophy wild-type protein ALDP.  
 KW Adrenoleucodystrophy; gene therapy; retroviral vector M48;  
 KW adrenomyelopathy; membrane protein; long chain fatty acid oxidation.  
 OS Homo sapiens.  
 PN W09621733-A2.  
 PD 18-JUL-1996.  
 PR 12-JAN-1996; F00059.  
 PR 13-JAN-1995; FR-000376.  
 PA (INRM ) INST NAT SANTE & RECH MEDICALE.  
 PI Aubourg P, Mandel JL, Mosser J, Sarde CO;  
 DR WPI: 96-342286/34.  
 DR N-PSDB; T39335.  
 PT Recombinant viral vector contg. DNA for correcting  
 PT adrenoleuco-dystrophy - and immunologically tolerable cells contg.  
 PT this vector, useful in gene therapy  
 PS Example; Fig 7; 36pp; French.  
 CC A 2.43 kb SpeI-EcoRI fragment of human wild-type ALD  
 CC (adrenoleucodystrophy) cDNA was inserted into retroviral vector M48  
 CC so that it was under control of the mouse phosphoglycerate kinase  
 CC (PGK) promoter. The resulting vector was co-transfected with vector  
 CC PGK-neomycin into amphotropic packaging cell line psiCRIP. Neomycin  
 CC (G418)-resistant transfectant cells which were also positive with  
 CC anti-ALD antibodies were incubated with skin fibroblasts of an ALD  
 CC patient. After two rounds of infection, about 70% of cells were  
 CC expressing normal ALD protein (ALDP) and integration of M48-ALD was  
 CC confirmed by Southern blotting. The present sequence is that of  
 CC ALDP.  
 SQ Sequence 745 AA;

Query Match 65.7%; Score 44; DB 20; Length 745;  
 Best Local Similarity 60.0%; Pred. No. 1.27e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 qapageptqe 56  
 QY 1 DGPTGEPOQE 10  
 :|||:|:

RESULT 10  
 ID R76110 standard; Protein; 745 AA.  
 AC R76110; 1995 (first entry)  
 DT 16-NOV-1995  
 DE Human ALD.  
 KW ALD gene; adrenoleukodystrophy; adrenomyeloneuropathy; AMN;  
 KW diagnosis; gene therapy.  
 OS Homo sapiens.  
 PN CA2108606-A.  
 PD 16-APR-1995.  
 PR 15-OCT-1993; 108606.  
 PR 15-OCT-1993; CA-108606.  
 PA (INRM ) INST NAT SANTE & RECH MEDICALE.  
 PI Aubourg P, Mandel J, Mosser J, Sarde C;  
 DR WPI: 95-215721/29.  
 DR N-PSDB; Q94048.  
 PT New nucleic acid responsible for adreno-leuco-dystrophy - related  
 PT probes, proteins and antibodies, useful for diagnosis and treatment  
 PS Claim 16; Fig.2A-B; 39pp; English.  
 CC Probes corresp. to breakpoints in the red pigment gene of an AMN  
 CC patient were used to isolate clones from an Xq28 cosmid library.  
 CC Following hybridizations and nested PCR, a HeLa cell cDNA library

CC was screened, and a complete ALD cDNA fragment (Q94048) encoding the protein given in R94048 was obt'd. Transformation of hematopoietic cells with ALD-encoding sequences in vivo will allow therapy of ALD or AMN.

SQ Sequence 745 AA;

Query Match 65.78; Score 44; DB 13; Length 745;  
Best Local Similarity 60.0%; Pred. No. 1.27e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 gapageptqe 56  
:::|||||

QY 1 DGPTGEPOQE 10

RESULT 11

ID W43039 standard; Protein; 943 AA.

AC W43039;

DT 12-MAY-1998 (first entry)

DE A NADPH oxidase derived from Arabidopsis thaliana.

KW NADPH oxidase; development; agricultural chemical; growth control;

KW transformation; plant; evaluation.

OS Arabidopsis thaliana.

FH Key Location/Qualifiers

FT Misc\_difference 161 /note= "not specified"

FT Misc\_difference 165 /note= "not specified"

FT

PN J10033176-A.

PD 10-FEB-1998.

PF 23-JUL-1996; 193220.

PR 23-JUL-1996; JP-193220.

PA (SUMO ) SUMITOMO CHEM CO LTD.

DR WPI: 98-172095/16.

DR N-PSDB; V03045.

PT NADPH oxidase derived from plants - useful for control of growth of transformed plants and for evaluation of agricultural chemicals

PS Claim 1; Pages 7-9; 11pp; Japanese.

CC This presents sequence represents a NADPH oxidase. The cDNA encoding this protein sequence was isolated from a cDNA library of Arabidopsis thaliana. The NADPH DNA and protein can be used in the development of agricultural chemicals. They allow for the control of growth of transformed plants for smooth evaluation of agricultural chemicals.

SQ Sequence 943 AA;

Query Match 64.28; Score 43; DB 28; Length 943;  
Best Local Similarity 60.0%; Pred. No. 1.63e+02;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 732 dgpygagagd 741  
|||||

QY 1 DGPTGEPOQE 10

RESULT 12

ID P93285 standard; Protein; 3080 AA.

AC P93285;

DT 06-APR-1990 (first entry)

DE Sequence of clone HIV-2 SBL/ISY.

KW HIV-2; proviral clone HIV-2 SBL/ISY.

OS Human immunodeficiency virus 2.

PN US7331212-A.

PD 29-AUG-1989.

PF 31-MAR-1989; 331212.

PR 31-MAR-1989; US-331212.

PA (USSH) US Dept. Health and Human Services.

PI Franchini G, Wong-Staal F, Gallo R;

DR WPI: 89-339698/46.

DR N-PSDB; N92119.

PT Complete human immunodeficiency type 2 proviral clone - used to generate animal model for function studies of HIV genes in vivo.

PS Disclosure; Fig. 5; 43pp; English.

CC The protein is encoded by the third reading frame of HIV-2 SBL/ISY, a

CC proviral clone of HIV-2.

SQ Sequence 3080 AA;

Query Match 64.28; Score 43; DB 1; Length 3080;  
Best Local Similarity 75.0%; Pred. No. 1.63e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 2799 phgepqqq 2806  
|||||

QY 3 PTGEPQOE 10

RESULT 13

ID P81771 standard; Protein; 3210 AA.

AC P81771;

DT 07-NOV-1990 (first entry)

DE Deduced sequence encoded by bottom reading frame of cDNA clone HIV-2 SBL/ISY of HIV related retrovirus strain

KW HIV vaccine; HIV strain SBL-6669-85.

OS Human immunodeficiency virus.

PN W08080449-A.

PD 03-NOV-1988.

PF 28-APR-1988; SE0218.

PR 28-APR-1987; SE-001765.

PA (SBLs-) SBL Statens Bekteri.

PI Albert J, Biberfeld G, Fenyo EM, Norrby E;

DR WPI: 88-322769/45.

DR N-PSDB; n80890.

PT HIV related human retro-virus strain - used for obtaining antigens for assays and vaccines and for prodn. of antibodies for assays

PS Claim 9; Fig 4; 28pp; English.

CC Synthetically produced proteins and peptides, characterised in that the AA sequence is derived from the primary nucleotide sequence of HIV-2 SBL/ISY or a part thereof, or a degenerate thereof are claimed.

CC HIV-2 SBL/ISY represents the complete genome of the virus SBL-6669 (-SBL-6669-85). The proviral DNA was obt'd. from a genomic library constructed from DNA of HUT-78 cells infected with SBL-6669-85 using the lambda-phage vector EMBL-3. SBL-6669-85 was isolated from lymphocytes of a West African woman. Protection is requested for the entire genome disclosed in n80890 and for parts thereof, and corresp. to various genes such as the gag gene (corresp. to nucleotides 547 to 2106), the pol gene (nucleotides 1827-4931) and the env gene (nucleotides 6144 to 8682), the corresp. AA sequences and parts thereof and various products derived therefrom, or use thereof, such as clones prep'd. by recombinant vector method, HIV test devices and methods. X corresponds to the translation of a stop codon.

SQ Sequence 3210 AA;

Query Match 64.28; Score 43; DB 1; Length 3210;  
Best Local Similarity 75.0%; Pred. No. 1.63e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 2906 phgepqqq 2913  
|||||

QY 3 PTGEPQOE 10

RESULT 14

ID R06684 standard; Protein; 17 AA.

AC R06684;

DT 11-JAN-1991 (first entry)

DE Human papilloma virus 16 E7-protein DNA sequence (V).

KW Immunogenic region; Human Papilloma Virus; Hpv16 E7; diagnosis;

KW antibodies; vaccines.

OS Human Papilloma Virus 16 E7.

PN EP-386734-A.

PD 12-SEP-1990.

PF 07-MAR-1990; 104353.

PR 10-MAR-1989; DE-907721.

PA (BEHW ) BEHRINGER AG.

PI Bartsch D, Gissmann L, Muller M;

DR WPI: 90-276785/37.

DR P-PSDB; R06684.  
PT New immunogenic regions of Papilloma virus 16 E7 protein - useful  
PS in vaccines and for diagnosis, and new derived antibodies  
PS Example 2; Page 3; 3pp; German.  
CC The sequence extends downstream from nucleotide 667 of a HPV16 DNA  
CC expression bank. HPV16 DNA fragments of ca. 100 bp were blunt-end  
CC ligated into PvuII-cleaved phage vector fd-tet-J6. Recombinant  
CC phages were plated on E. coli K91 and replicated on nitrocellulose  
CC membranes probed with specific sera. 200 recombinants reacted and  
CC 30 of them were sequenced. This sequence was identified from 5  
CC overlapping clones. The sequence encodes immunogenic regions which  
CC are useful in vaccines, to detect specific antibodies against HPV16  
CC E7-protein. Antibodies are also useful diagnostically.  
CC See also Q03881.  
SQ Sequence 17 AA;

Query Match 62.7%; Score 42; DB 2; Length 17;  
Best Local Similarity 40.0%; Pred. No. 2.09e+02;  
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 4 dgpaggaeqd 13  
|||:|:::  
QY 1 DGPtGEFQQE 10

RESULT 15  
ID R14295 standard; Protein; 17 AA.  
AC R14295;  
DT 02-JAN-1992 (first entry)  
DE Seroreactive epitope #1 of HPV16 genomic region E7-107.  
KW HPV16-dependent human disease; E4; E6; L1.  
OS Synthetic.  
PN EP-451550-A.  
PD 16-OCT-1991.  
PF 19-MAR-1991; 104197.  
PR 20-MAR-1990; EP-105222.  
PR 19-MAR-1991; EP-104197.  
PA (BEHW ) BEHRINGWERKE AG.  
PI Muller M, Gissmann L;  
DR WFI; 91-304643/42.  
PT Sero-active epitope(s) of human papilloma-virus 16 proteins - for  
PT use vaccines and diagnosis  
PS Claim 4; Page 11; 15pp; English.  
CC This is one of two seroreactive epitopes derived from the 107  
CC region of HPV16 protein E7. Peptides containing this epitope are  
CC also claimed and can be used to generate antibodies to HPV.  
CC See R14287-R14302 and Q14168-Q14171.  
SQ Sequence 17 AA;

Query Match 62.7%; Score 42; DB 3; Length 17;  
Best Local Similarity 40.0%; Pred. No. 2.09e+02;  
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 4 dgpaggaeqd 13  
|||:|:::  
QY 1 DGPtGEFQQE 10

Search completed: Thu May 13 15:36:03 1999  
Job time : 17 secs.



\*\*\*\*\*

WIREH

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu May 13 15:40:37 1999; MasPar time 2.71 Seconds  
Tabular output not generated. 82.989 Million cell updates/sec

Title: >US-09-040-485-8  
Description: (3-8) from US09040485.pep (2 of 2)  
Perfect Score: 41  
Sequence: 1 PTGEPQ 6

Scoring table: PAM 150  
Gap 15

Searched: 116738 seqs, 37463448 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir58  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 17.462; Variance 17.910; scale 0.975

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description                        | Pred. No. |
|------------|-------|---------------|--------|----|------------------------------------|-----------|
| 1          | 41    | 100.0         | 115    | 2  | A23925 proline-rich phosphoprotein | 1.72e+00  |
| 2          | 41    | 100.0         | 757    | 2  | I38423 aspartyl beta-hydroxylase   | 1.72e+00  |
| 3          | 38    | 92.7          | 232    | 1  | MMBBI3 25.5k membrane protein      | 9.96e+00  |
| 4          | 38    | 92.7          | 484    | 1  | SYR2ET glutamate--tRNA ligas       | 9.96e+00  |
| 5          | 37    | 90.2          | 325    | 1  | B48561 inner capsid protein        | 1.75e+01  |
| 6          | 37    | 90.2          | 325    | 1  | A48561 inner capsid protein        | 1.75e+01  |
| 7          | 37    | 90.2          | 328    | 1  | JQ1875 inner capsid protein        | 1.75e+01  |
| 8          | 37    | 90.2          | 328    | 1  | VPXRC3 minor inner core prot       | 1.75e+01  |
| 9          | 37    | 90.2          | 369    | 2  | A36909 C4-dicarboxylase-bind       | 1.75e+01  |
| 10         | 37    | 90.2          | 370    | 2  | A36909 glutamyl-tRNA synthet       | 1.75e+01  |
| 11         | 37    | 90.2          | 468    | 2  | S21172 glutamate--tRNA ligas       | 1.75e+01  |
| 12         | 37    | 90.2          | 494    | 2  | S76516 integral membrane pro       | 1.75e+01  |
| 13         | 37    | 90.2          | 505    | 2  | JCS208 glutamate--tRNA ligas       | 1.75e+01  |
| 14         | 37    | 90.2          | 506    | 2  | F71513 probable glutamyl-trn       | 1.75e+01  |
| 15         | 36    | 87.8          | 142    | 1  | SKPSXT secretion protein xcp       | 3.05e+01  |
| 16         | 36    | 87.8          | 612    | 2  | G69649 GTP-binding protein i       | 3.05e+01  |
| 17         | 35    | 87.8          | 770    | 2  | S60876 cellobiose oxidase (E       | 3.05e+01  |
| 18         | 35    | 85.4          | 231    | 2  | S76204 hypothetical factor         | 5.26e+01  |
| 19         | 35    | 85.4          | 235    | 2  | S23249 transcription protein       | 5.26e+01  |
| 20         | 35    | 85.4          | 251    | 1  | PIHUPF salivary proline-rich       | 5.26e+01  |
| 21         | 35    | 85.4          | 303    | 2  | C70808 hypothetical protein        | 5.26e+01  |
| 22         | 35    | 85.4          | 324    | 2  | SL1497 CAMP-binding protein        | 5.26e+01  |
| 23         | 35    | 85.4          | 333    | 2  | SL1484 CABP1 protein - slime       | 5.26e+01  |

|    |    |      |      |   |                               |          |
|----|----|------|------|---|-------------------------------|----------|
| 24 | 35 | 85.4 | 337  | 2 | A37760 UDPglucose 4-epimeras  | 5.26e+01 |
| 25 | 35 | 85.4 | 392  | 1 | PIHUB6 salivary proline-rich  | 5.26e+01 |
| 26 | 35 | 85.4 | 441  | 2 | S74800 dihydroorotase (EC 3.  | 5.26e+01 |
| 27 | 35 | 85.4 | 512  | 2 | SL9354 cell fusion protein F  | 5.26e+01 |
| 28 | 35 | 85.4 | 535  | 2 | P70782 probable oxidoreducta  | 5.26e+01 |
| 29 | 35 | 85.4 | 545  | 2 | A39193 cytochrome c3 precu    | 5.26e+01 |
| 30 | 35 | 85.4 | 562  | 2 | A41707 oligo-1,6-glucosidase  | 5.26e+01 |
| 31 | 35 | 85.4 | 600  | 2 | H70448 G-protein LepA - Aqu   | 5.26e+01 |
| 32 | 35 | 85.4 | 622  | 2 | I48733 protein kinase rck (E  | 5.26e+01 |
| 33 | 35 | 85.4 | 622  | 2 | A34711 kinase-related protei  | 5.26e+01 |
| 34 | 35 | 85.4 | 666  | 2 | B70803 hypothetical protein   | 5.26e+01 |
| 35 | 35 | 85.4 | 2336 | 2 | A45386 omega-conotoxin-sensi  | 5.26e+01 |
| 36 | 35 | 85.4 | 3712 | 2 | S18253 laminin alpha-1 chain  | 5.26e+01 |
| 37 | 34 | 82.9 | 90   | 3 | JCS931 high mobility group I  | 8.97e+01 |
| 38 | 34 | 82.9 | 96   | 3 | JCS933 high mobility group I  | 8.97e+01 |
| 39 | 34 | 82.9 | 105  | 3 | JCS932 high mobility group I  | 8.97e+01 |
| 40 | 34 | 82.9 | 436  | 2 | S69821 NDP-sugar dehydrogena  | 8.97e+01 |
| 41 | 34 | 82.9 | 597  | 2 | G71373 probable methyl-accept | 8.97e+01 |
| 42 | 34 | 82.9 | 899  | 1 | A43969 core protein VP3 - ep  | 8.97e+01 |
| 43 | 34 | 82.9 | 901  | 2 | S07419 core protein P3 - blu  | 8.97e+01 |
| 44 | 34 | 82.9 | 1838 | 1 | CGHUIV collagen alpha 1(V) c  | 8.97e+01 |
| 45 | 34 | 82.9 | 1843 | 2 | SL8803 collagen alpha 1(V) c  | 8.97e+01 |

## ALIGNMENTS

RESULT 1  
ENTRY proline-rich phosphoprotein - crab-eating macaque  
TITLE #formal\_name Macaca fascicularis #common\_name crab-eating macaque  
ORGANISM  
DATE 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 25-Oct-1996  
ACCESSIONS A23925 #type complete  
REFERENCE A23925 proline-rich phosphoprotein - crab-eating macaque  
#authors Oppenheim, F.G.; Offner, G.D.; Troxler, R.F.  
#journal J Biol Chem 263:10611-10628  
#title Amino acid sequence of a proline-rich phosphoglycoprotein from parotid secretion of the subhuman primate Macaca fascicularis.  
#cross-references MUID:85289254  
#accession A23925  
#molecule\_type protein  
#residues 1-115 #label OPP  
CLASSIFICATION #superfamily proline-rich protein  
KEYWORDS phosphoprotein  
SUMMARY #length 115 #molecular-weight 12255 #checksum 2737

Query Match 100.0%; Score 41; DB 2; Length 115;  
Best Local Similarity 100.0%; Pred. No. 1.72e+00;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 68 PTGEPQ 73

QY 3 PTGEPQ 8

RESULT 2

ENTRY I38423 #type complete  
TITLE aspartyl beta-hydroxylase - human  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 10-Jul-1998  
ACCESSIONS I38423  
REFERENCE I38423  
#authors Koriath, F.; Gieffers, C.; Frey, J.  
#journal Gene (1994) 150:395-399  
#title Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase.  
#cross-references MUID:95121937  
#accession I38423  
#status Preliminary; translated from GB/EMBL/DBJ  
#molecule\_type mRNA

```
##residues 1-757 ##label RES
##cross-references EMBL:U03109; NID:g458031; PID:g458032
CLASSIFICATION #superfamily peptide-aspartate beta-dioxygenase;
tetrairicopeptide repeat homology
FEATURE
54-75 #domain transmembrane #status predicted #label TRM
SUMMARY #length 757 #molecular-weight 85498 #checksum 2143

Query Match 100.0%; Score 41; DB 2; Length 757;
Best Local Similarity 100.0%; Pred. No. 1.72e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 175 PTGEPQ 180
|||||
Qy 3 PTGEPQ 8

RESULT 3
ENTRY MMBEI3 #type complete
TITLE 25.5K membrane protein - ictalurid herpesvirus 1 (strain
aurn 1)
ORGANISM #formal_name ictalurid herpesvirus 1
#note host Ictalurus punctatus (channel catfish)
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
05-Sep-1997
ACCESSIONS I36786
REFERENCE I36804
#authors Davison, A.J.
#submission submitted to GenBank, January 1992
#description Channel catfish virus: a new type of herpesvirus.
#accession I36786
#molecule_type DNA
##residues 1-232 ##label DAV
##cross-references GB:M75136; NID:g331209; PID:g331218
REFERENCE A39447
#authors Davison, A.J.
#journal Virology (1992) 186:9-14
#title Channel catfish virus: a new type of herpesvirus.
#cross-references MUID:92087490
#contents annotation
#note neither amino acid nor nucleotide sequence is given
GENETICS
#gene 8
CLASSIFICATION #superfamily ictalurid herpesvirus 25.5K membrane protein
transmembrane protein
FEATURE
166-182 #domain transmembrane #status predicted #label TW1\
195-211 #domain transmembrane #status predicted #label TW2
SUMMARY #length 232 #molecular-weight 25466 #checksum 8650

Query Match 92.7%; Score 38; DB 1; Length 232;
Best Local Similarity 83.3%; Pred. No. 9.96e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 90 PTGEPH 95
|||||
Qy 3 PTGEPQ 8

RESULT 4
ENTRY SYRZET #type complete
TITLE glutamate--tRNA ligase (EC 6.1.1.17) - Rhizobium melliloti
ALTERNATE_NAMES glutamyl-tRNA synthetase
ORGANISM #formal_name Rhizobium melliloti
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
05-Sep-1997
ACCESSIONS A32888
REFERENCE A32888
#authors Laberge, S.; Gagnon, Y.; Bordeleau, L.M.; Lapointe, J.
#journal J. Bacteriol. (1989) 171:3926-3932
#title Cloning and sequencing of the gltx gene, encoding the
glutamyl-tRNA synthetase of Rhizobium melliloti A2.
#cross-references MUID:89291743

#accession A32888
#molecule_type DNA
#residues 1-325 ##label LAB
#cross-references GB:M27221; NID:g341649; PID:g717082
#experimental_source strain A2
CLASSIFICATION #superfamily glutamate--tRNA ligase
KEYWORDS aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis
FEATURE
256 #binding_site ATP (Lys) #status predicted
SUMMARY #length 484 #molecular-weight 54228 #checksum 7829

Query Match 92.7%; Score 38; DB 1; Length 484;
Best Local Similarity 83.3%; Pred. No. 9.96e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 14 PTGEPH 19
|||||
Qy 3 PTGEPQ 8

RESULT 5
ENTRY B48561 #type complete
TITLE inner capsid protein VP6 - bluetongue virus (serotype 10,
strain USA)
ORGANISM #formal_name bluetongue virus
DATE 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change
20-Mar-1998
ACCESSIONS B48561
REFERENCE A48561
#authors Hwang, G.Y.; Chiou, J.F.; Yang, Y.Y.; Li, J.K.
#journal Virus Res. (1992) 24:315-323
#title Comparative sequence analyses of the cognate structural
protein VP6 genes of five US bluetongue viruses.
#cross-references MUID:93033709
#accession B48561
#molecule_type genomic RNA
#residues 1-325 ##label HWA
#cross-references GB:L08669; NID:g210842; PID:g210843
#note sequence extracted from NCBI backbone (NCBIP:115445)
GENETICS
#map_position segment 9
CLASSIFICATION #superfamily bluetongue virus VP6 protein
capsid protein
KEYWORDS #length 325 #molecular-weight 35163 #checksum 4288
SUMMARY

Query Match 90.2%; Score 37; DB 1; Length 325;
Best Local Similarity 66.7%; Pred. No. 1.75e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 285 PTGDPH 290
|||||
Qy 3 PTGEPQ 8

RESULT 6
ENTRY A48561 #type complete
TITLE inner capsid protein VP6 - bluetongue virus (serotype 11,
strain USA)
ORGANISM #formal_name bluetongue virus
DATE 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change
17-Feb-1994
ACCESSIONS A48561
REFERENCE A48561
#authors Hwang, G.Y.; Chiou, J.F.; Yang, Y.Y.; Li, J.K.
#journal Virus Res. (1992) 24:315-323
#title Comparative sequence analyses of the cognate structural
protein VP6 genes of five US bluetongue viruses.
#cross-references MUID:93033709
#accession A48561
#molecule_type genomic RNA
#residues 1-325 ##label HWA
#note sequence extracted from NCBI backbone (NCBIP:115444)
```

## GENETICS

#map\_position segment 9  
 CLASSIFICATION #superfamily bluetongue virus VP6 protein  
 KEYWORDS capsid protein  
 SUMMARY #length 325 #molecular-weight 33554 #checksum 2927

Query Match 90.2%; Score 37; DB 1; Length 325;  
 Best Local Similarity 66.7%; Pred. No. 1.75e+01;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 285 PTGDPH 290  
 |||:  
 Qy 3 PTGEPQ 8

## RESULT 7

ENTRY JQ1875 #type complete  
 TITLE inner capsid protein VP6 - bluetongue virus (serotype 1, strain South Africa)  
 ALTERNATE\_NAMES VP6 protein  
 ORGANISM #formal\_name bluetongue virus  
 DATE 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 05-Sep-1997

## ACCESSIONS

REFERENCE JQ1875  
 #authors Wade-Evans, A.M.; Mertens, P.P.C.; Belsham, G.J.  
 #journal J. Gen. Virol. (1992) 73:3023-3026  
 #title Sequence of genome segment 9 of bluetongue virus (serotype 1, South Africa) and expression analysis demonstrating that different forms of VP6 are derived from initiation of protein synthesis at two distinct sites.

#accession JQ1875  
 #molecule\_type genomic RNA  
 #residues 1-328 #label WAD  
 #cross-references GB:D10905; NID:G221080; PID:d1002188; PID:g221081

## GENETICS

#map\_position segment 9  
 CLASSIFICATION #superfamily bluetongue virus VP6 protein  
 KEYWORDS capsid protein  
 SUMMARY #length 328 #molecular-weight 35928 #checksum 4514

Query Match 90.2%; Score 37; DB 1; Length 328;  
 Best Local Similarity 66.7%; Pred. No. 1.75e+01;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294  
 |||:  
 Qy 3 PTGEPQ 8

## RESULT 8

ENTRY VPXRC3 #type complete  
 TITLE minor inner core protein VP6 - bluetongue virus (serotype 10, American isolate)  
 ALTERNATE\_NAMES inner capsid protein VP6; VP6 protein  
 ORGANISM #formal\_name bluetongue virus  
 DATE 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 10-Nov-1995

ACCESSIONS B32400; S10542  
 REFERENCE A32400

#authors Fukusho, A.; Yu, Y.; Yamaguchi, S.; Roy, P.  
 #journal J. Gen. Virol. (1989) 70:1677-1689  
 #title Completion of the sequence of bluetongue virus serotype 10 by the characterization of a structural protein, VP6, and a non-structural protein, NS2.

#cross-references MUID:89293076

#accession B32400  
 #molecule\_type genomic RNA  
 #residues 1-328 #label FUK

## REFERENCE

S10534  
 #authors Roy, P.; Marshall, J. J. A.; French, T. J.  
 #journal Curr. Top. Microbiol. Immunol. (1990) 162:43-87  
 #title Structure of the bluetongue virus genome and its encoded

proteins.

#cross-references MUID:90345726  
 #accession S10542  
 #status preliminary  
 #molecule\_type genomic RNA  
 #residues 1-328 #label CUR

## GENETICS

#map\_position segment 9  
 CLASSIFICATION #superfamily bluetongue virus VP6 protein  
 KEYWORDS capsid protein  
 SUMMARY #length 328 #molecular-weight 35749 #checksum 5072

Query Match 90.2%; Score 37; DB 1; Length 328;  
 Best Local Similarity 66.7%; Pred. No. 1.75e+01;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294  
 |||:  
 Qy 3 PTGEPQ 8

## RESULT 9

ENTRY S77299 #type complete  
 TITLE C4-dicarboxylase-binding protein - Synecocystis sp. (strain PCC 6803)  
 ALTERNATE\_NAMES protein sll1314  
 ORGANISM #formal\_name Synecocystis sp.  
 #variety PCC 6803  
 DATE 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Aug-1998

ACCESSIONS S77299  
 REFERENCE S74322

#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.  
 #journal DNA Res. (1996) 3:109-136  
 #title Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.

#cross-references MUID:97061201

#accession S77299  
 #status nucleic acid sequence not shown; translation not shown  
 #molecule\_type DNA  
 #residues 1-369 #label KAN  
 #cross-references EMBL:D90907; GB:AB001339; NID:g1652618; PID:d1018366; PID:g1652713

##note the nucleotide sequence was submitted to the EMBL Data Library, June 1996

## GENETICS

#gene dctp  
 SUMMARY #length 369 #molecular-weight 41164 #checksum 6843

Query Match 90.2%; Score 37; DB 2; Length 369;  
 Best Local Similarity 66.7%; Pred. No. 1.75e+01;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 99 PAGDPQ 104  
 |||:  
 Qy 3 PTGEPQ 8

## RESULT 10

ENTRY A36909 #type fragment  
 TITLE glutamyl-tRNA synthetase homolog gltX - Chlamydia psittaci (fragment)

ORGANISM #formal\_name Chlamydia psittaci  
 DATE 07-Apr-1994 #sequence\_revision 18-Nov-1994 #text\_change 20-Mar-1998  
 ACCESSIONS A36909

```

REFERENCE      A36909
#authors      Wichlan, D.G.; Hatch, T.P.
#journal      J. Bacteriol. (1993) 175:2936-2942
#title        Identification of an early-stage gene of Chlamydia psittaci
              68C.
#cross-references MUID:93259937
#contents     68C
#accession    A36909
#status       preliminary
#molecule_type DNA
#residues     1-370 ##label WIC
##cross-references GB:113598; GB:M73485; NID:g289832; PID:g289833
#note         sequence extracted from NCBI backbone (NCBIN:131863,
              NCBIP:131864)
CLASSIFICATION #superfamily glutamate--trna ligase
SUMMARY        #length 370 #checksum 8346

Query Match      90.2%; Score 37; DB 2; Length 370;
Best Local Similarity 66.7%; Pred. No. 1.75e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 16 PTGDPH 21
|||:|
Qy 3 PTGEPQ 8

RESULT 11
ENTRY      S21172      #type complete
TITLE      glutamate--trna ligase (EC 6.1.1.17) - Thermus aquaticus
ORGANISM    22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change
DATE        24-Apr-1998
ACCESSIONS S21172; S21236
REFERENCE   S21172
#authors    Nureki, O.; Suzuki, K.; Hara-Yokoyama, M.; Kohno, T.;
              Matsuzawa, H.; Ohta, T.; Shimizu, T.; Morikawa, K.;
              Miyazawa, T.; Yokoyama, S.
#journal    Eur. J. Biochem. (1992) 204:465-472
#title      GlutamyI-trna synthetase from Thermus thermophilus HB8.
              Molecular cloning of the gltX gene and crystallization of
              the overproduced protein.
#cross-references MUID:92174899
#accession  S21172
#molecule_type DNA
#residues   1-468 ##label NUR
##cross-references EMBL:X64557; NID:g48241; PID:g48242
#note       the sequence from Fig. 4 is inconsistent with that from
              Fig. 3 in having 67-Thr
#note       the source is designated as Thermus thermophilus
#accession  S21236
#molecule_type DNA
#residues   1-40 ##label NU2
#note       the source is designated as Thermus thermophilus
CLASSIFICATION #superfamily glutamate--trna ligase
KEYWORDS      aminoacyl-trna synthetase; ligase; protein biosynthesis
SUMMARY        #length 468 #molecular-weight 53901 #checksum 2454

Query Match      90.2%; Score 37; DB 2; Length 468;
Best Local Similarity 66.7%; Pred. No. 1.75e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 10 PTGDPH 15
|||:|
Qy 3 PTGEPQ 8

RESULT 12
ENTRY      S76516      #type complete
TITLE      integral membrane protein - Synecocystis sp. (strain PCC
              6803)
ALTERNATE_NAMES protein slr0642
ORGANISM      #formal_name Synecocystis sp.
#variety      PCC 6803

```

```

DATE          25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
              21-Aug-1998
ACCESSIONS    S76516
REFERENCE     S74322
#authors      Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
              Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugiyama, M.;
              Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
              Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
              S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
              Yasuda, M.; Tabata, S.
#journal      DNA Res. (1996) 3:109-136
#title        Sequence analysis of the genome of the unicellular
              cyanobacterium Synecocystis sp. PCC6803. II. Sequence
              determination of the entire genome and assignment of
              potential protein-coding regions.
#cross-references MUID:97061201
#accession    S76516
#status       nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues     1-494 ##label KAN
##cross-references EMBL:D64002; GB:AB001339; NID:gl001612; PID:dl011013;
              PID:gl001631
#note         the nucleotide sequence was submitted to the EMBL Data
              Library, June 1996
GENETICS      #start_codon GTG
KEYWORDS      membrane protein
SUMMARY        #length 494 #molecular-weight 53585 #checksum 7038

Query Match      90.2%; Score 37; DB 2; Length 494;
Best Local Similarity 66.7%; Pred. No. 1.75e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 437 PAGDPQ 442
|||:|
Qy 3 PTGEPQ 8

RESULT 13
ENTRY      JC5208      #type complete
TITLE      glutamate--trna ligase (EC 6.1.1.17) - Chlamydia psittaci
ALTERNATE_NAMES glutamyl-trna synthetase
ORGANISM        #formal_name Chlamydia psittaci
DATE            20-Feb-1997 #sequence_revision 27-Feb-1997 #text_change
              14-Nov-1997
ACCESSIONS    JC5208
REFERENCE     JC5208
#authors      Hsia, R.; Bavoil, P.M.
#journal      Gene (1996) 176:163-169
#title        Homologs of Escherichia coli recJ, gltX and of a putative
              'early' gene of avian Chlamydia psittaci are located
              upstream of the 'late' omp2 locus of Chlamydia psittaci
              strain guinea pig inclusion conjunctivitis.
#accession    JC5208
#status       preliminary
#molecule_type DNA
#residues     1-505 ##label HSI
##cross-references GB:U41759; NID:g1783376; PID:g1783380
#experimental_source strain GPIC
CLASSIFICATION #superfamily glutamate--trna ligase
KEYWORDS      aminoacyl-trna synthetase; ligase; protein biosynthesis
FEATURE        9-11
              322
              66
SUMMARY        #length 505 #molecular-weight 58544 #checksum 9458

Query Match      90.2%; Score 37; DB 2; Length 505;
Best Local Similarity 66.7%; Pred. No. 1.75e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 14 PTGDPH 19
|||:|

```

QY 3 PTGEPQ 8

## RESULT 14

ENTRY F71513 #type complete  
TITLE Probable glutamyl-tRNA synthetase - Chlamydia trachomatis  
(serotype D, strain UW3/Cx)  
ORGANISM #formal\_name Chlamydia trachomatis  
DATE 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change  
13-Sep-1998  
ACCESSIONS F71513  
REFERENCE A71460  
#authors Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe,  
R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov,  
R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.  
#submission submitted to GenBank, May 1998  
#description Genome sequence of an obligate intracellular pathogen of  
humans: Chlamydia trachomatis.  
#accession F71513  
##status preliminary  
##molecule\_type DNA  
##residues 1-506 ##label ARN  
##cross-references GB:AE001318; GB:AE001273; NID:g3328875; PID:g3328877  
##experimental\_source serotype D, strain UW-3/Cx

## GENETICS

#gene gltx  
SUMMARY #length 506 #molecular-weight 58510 #checksum 7203

Query Match 90.2%; Score 37; DB 2; Length 506;  
Best Local Similarity 66.7%; Pred. NO. 1.75e+01;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 14 PTGDPH 19

||||

QY 3 PTGEPQ 8

## RESULT 15

ENTRY SKPSXT #type complete  
TITLE secretion protein xcpT - Pseudomonas aeruginosa  
ORGANISM #formal\_name Pseudomonas aeruginosa  
DATE 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change  
05-Sep-1997  
ACCESSIONS S25386; S22731  
REFERENCE S25384  
#authors Bally, M.; Filloux, A.; Akrim, M.; Ball, G.; Lazdunski, A.;  
Tomassen, J.  
#journal Mol. Microbiol. (1992) 6:1121-1131  
#title Protein secretion in Pseudomonas aeruginosa: characterization  
of seven xcp genes and processing of secretory apparatus  
components by prepilin peptidase.  
#accession S25386  
##molecule\_type DNA  
##residues 1-142 ##label BAL  
##cross-references EMBL:X62666; NID:g45433; PID:g581440  
GENETICS  
#gene xcpT  
#start\_codon TTG  
CLASSIFICATION #superfamily secretion protein xcpT  
SUMMARY #length 142 #molecular-weight 15449 #checksum 2410

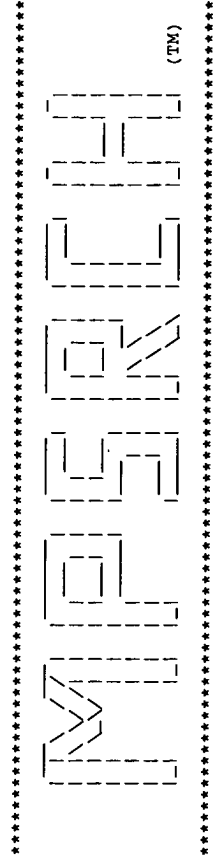
Query Match 87.8%; Score 36; DB 1; Length 142;  
Best Local Similarity 83.3%; Pred. NO. 3.05e+01;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 79 PTGNPQ 84

||||

QY 3 PTGEPQ 8

Search completed: Thu May 13 15:40:47 1999  
Job time : 10 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPerch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 13 15:36:54 1999; MasPar time 2.16 Seconds  
Tabular output not generated. 124.372 Million cell updates/sec

Title: >US-09-040-485-8  
Description: (1-10) from US09040485.pep (1 of 2)  
Perfect Score: 67  
Sequence: 1 DGPTGEPOQE 10

Scoring table: PAM 150  
Gap 15

Searched: 74019 seqs, 26840295 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot35  
1:swissprot

Statistics: Mean 21.176; Variance 23.637; scale 0.896

Préd. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES  |       |               |        |    |             |                         |          |
|------------|-------|---------------|--------|----|-------------|-------------------------|----------|
| Result No. | Score | Query Match % | Length | ID | Description | Pred. No.               |          |
| 1          | 67    | 100.0         | 757    | 1  | ASPH_HUMAN  | ASPARTYL/ASPARAGINYL B  | 1.38e-05 |
| 2          | 50    | 74.6          | 232    | 1  | VG08_HSVI1  | HYPOTHETICAL GENE 8 ME  | 3.41e-01 |
| 3          | 47    | 70.1          | 433    | 1  | TRBI_AGR76  | CONJUGAL TRANSFER PROT  | 1.72e+00 |
| 4          | 47    | 70.1          | 1389   | 1  | TIMDROME    | TIMELESS PROTEIN        | 1.72e+00 |
| 5          | 47    | 70.1          | 3567   | 1  | ERY2_SACER  | ERYTHRONOLIDE SYNTHASE  | 1.72e+00 |
| 6          | 46    | 68.7          | 311    | 1  | SRY_MUSST   | SEX-DETERMINING REGION  | 2.90e+00 |
| 7          | 46    | 68.7          | 355    | 1  | SRY_MUSST   | SEX-DETERMINING REGION  | 2.90e+00 |
| 8          | 45    | 67.2          | 53     | 1  | CALL1_RABIT | COLLAGEN ALPHA 1(I) CH  | 4.85e+00 |
| 9          | 45    | 67.2          | 259    | 1  | YCBC_ECOLI  | HYPOTHETICAL 28.7 KD P  | 4.85e+00 |
| 10         | 45    | 67.2          | 1027   | 1  | P531_HUMAN  | P53-BINDING PROTEIN 53  | 4.85e+00 |
| 11         | 45    | 67.2          | 1418   | 1  | CALL2_HUMAN | PROCOLLAGEN ALPHA 1(II) | 4.85e+00 |
| 12         | 45    | 67.2          | 1596   | 1  | MAM1_DROME  | PROCOLLAGEN ALPHA 1(II) | 4.85e+00 |
| 13         | 44    | 65.7          | 261    | 1  | PRP2_MOUSE  | PROLINE-RICH PROTEIN M  | 8.07e+00 |
| 14         | 44    | 65.7          | 745    | 1  | ALD1_HUMAN  | ADRENOLEUKODYSTROPHY P  | 8.07e+00 |
| 15         | 44    | 65.7          | 755    | 1  | RRE1_HUMAN  | RAS-RESPONSIVE ELEMENT  | 8.07e+00 |
| 16         | 44    | 65.7          | 1388   | 1  | CALL1_HUMAN | COLLAGEN ALPHA 1(XV) C  | 8.07e+00 |
| 17         | 43    | 64.2          | 336    | 1  | CALL6_HUMAN | COLLAGEN ALPHA 1(VI) C  | 1.33e+01 |
| 18         | 43    | 64.2          | 449    | 1  | CMGA_BOVIN  | CHROMOGRANIN A PRECURS  | 1.33e+01 |
| 19         | 43    | 64.2          | 1670   | 1  | CA34_HUMAN  | PROCOLLAGEN ALPHA 3(IV) | 1.33e+01 |
| 20         | 43    | 64.2          | 1707   | 1  | CA24_MOUSE  | PROCOLLAGEN ALPHA 2(IV) | 1.33e+01 |
| 21         | 43    | 64.2          | 1841   | 1  | RPB1_ARATH  | DNA-DIRECTED RNA POLYM  | 1.33e+01 |
| 22         | 43    | 64.2          | 1860   | 1  | RPB0_ARATH  | DNA-DIRECTED RNA POLYM  | 1.33e+01 |
| 23         | 42    | 62.7          | 181    | 1  | HP28_RAT    | 28 KD HEAT- AND ACID-S  | 2.17e+01 |

| RESULT ID | ASPH_HUMAN   | STANDARD | PRT | 757 AA. |
|-----------|--|----------|-----|---------|
| AC        | Q12797   |          |     |         |
| DT        | 01-NOV-1997 (REL. 35, CREATED)   |          |     |         |
| DT        | 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  |          |     |         |
| DT        | 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  |          |     |         |
| DE        | ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-HYDROXYLASE) (ASP BETA-HYDROXYLASE) (PEPTIDE-ASPARTATE BETA-HYDROXYLASE) |          |     |         |
| DE        | DIOXYGENASE)   |          |     |         |
| GN        | ASPH.  |          |     |         |
| OS        | HOMO SAPIENS (HUMAN).  |          |     |         |
| OC        | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;   |          |     |         |
| OC        | EUTHERIA; PRIMATES.  |          |     |         |
| RN        | [1]  |          |     |         |
| RP        | SEQUENCE FROM N.A.   |          |     |         |
| RX        | MEDLINE: 95121937.   |          |     |         |
| RA        | KORIOTH F., GIEFFERS C., FREY J.;  |          |     |         |
| RL        | GENE 1501395-1591394   |          |     |         |

ALIGNMENTS

| RESULT ID | ASPH_HUMAN   | STANDARD | PRT | 757 AA. |
|-----------|--|----------|-----|---------|
| AC        | Q12797   |          |     |         |
| DT        | 01-NOV-1997 (REL. 35, CREATED)   |          |     |         |
| DT        | 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  |          |     |         |
| DT        | 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  |          |     |         |
| DE        | ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-HYDROXYLASE) (ASP BETA-HYDROXYLASE) (PEPTIDE-ASPARTATE BETA-HYDROXYLASE) |          |     |         |
| DE        | DIOXYGENASE)   |          |     |         |
| GN        | ASPH.  |          |     |         |
| OS        | HOMO SAPIENS (HUMAN).  |          |     |         |
| OC        | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;   |          |     |         |
| OC        | EUTHERIA; PRIMATES.  |          |     |         |
| RN        | [1]  |          |     |         |
| RP        | SEQUENCE FROM N.A.   |          |     |         |
| RX        | MEDLINE: 95121937.   |          |     |         |
| RA        | KORIOTH F., GIEFFERS C., FREY J.;  |          |     |         |
| RL        | GENE 1501395-1591394   |          |     |         |
| CC        | -1- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF PROTEINS.   |          |     |         |
| CC        | -1- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) -> PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).                      |          |     |         |
| CC        | -1- COFACTOR: IRON.  |          |     |         |
| CC        | -1- SUBUNIT: MONOMER (BY SIMILARITY).  |          |     |         |
| CC        | -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM.   |          |     |         |
| CC        | -1- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES TESTED.  |          |     |         |
| CC        | -1- PFM: MIGHT BE PROCESSED TO THE 56 KD (AA 274-757) OR 52 KD (AA 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY SIMILARITY).   |          |     |         |
| DR        | EMBL: U03109; E82591; ..   |          |     |         |
| DR        | MIM: 600582; ..  |          |     |         |
| KW        | Oxidoreductase; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR; ENDOPLASMIC RETICULUM.  |          |     |         |
| FT        | DOMAIN 1 54  |          |     |         |
| FT        | TRANSMEM 55 75   |          |     |         |
| FT        | DOMAIN 76 757  |          |     |         |
| FT        | DOMAIN 13 20   |          |     |         |
| FT        | DOMAIN 323 332   |          |     |         |
| FT        | POLY-SER.  |          |     |         |
| FT        | POLY-LYS.  |          |     |         |
| FT        | CARBOHYD 452   |          |     |         |
| FT        | POTENTIAL.   |          |     |         |
| FT        | POTENTIAL.   |          |     |         |
| SQ        | SEQUENCE 757 AA; 85498 MW; AE6AFC24 CRC32;   |          |     |         |

Query Match 100.0%; Score 67; DB 1; Length 757;  
Best Local Similarity 100.0%; Pred. No. 1.38e-05;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 173 DGTPGEPOQ 182  
|||||  
QY 1 DGTPGEPOQ 10

RESULT 2  
ID VG08\_HSV11 STANDARD; PRT; 232 AA.

AC Q00137;  
DT 01-DEC-1992 (REL. 24, CREATED)  
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL GENE 8 MEMBRANE PROTEIN.  
GN 8.

OS ICTALURID HERPESVIRUS 1 (CHANNEL CATFISH VIRUS) (CCV).

OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-AUBURN 1.

RX MEDLINE; 92087490.

RA DAVISON A.J.;

RL VIROLOGY 186:9-14(1992).

DR EMBL; M75136; G331295; -.

DR EMBL; M75136; G331218; -.

PIR; I36786; MMBE13

KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.

FT TRANSMEM 166 182 POTENTIAL.

FT TRANSMEM 195 211 POTENTIAL.

SQ SEQUENCE 232 AA; 25466 MW; AA90E31E CRC32;

Query Match 74.6%; Score 50; DB 1; Length 232;

Best Local Similarity 50.0%; Pred. No. 3.41e-01;

Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 88 ESPGEPHRD 97  
:|||||  
QY 1 DGTPGEPOQ 10

RESULT 3

ID TRB1\_AGR6 STANDARD; PRT; 433 AA.

AC P54917;

DT 01-OCT-1996 (REL. 34, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE CONJUGAL TRANSFER PROTEIN TRB1.

GN TRB1.

OS AGROBACTERIUM TUMEFACIENS.

OG PLASMID PTIA6C.

OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;

OC RHIZOBIACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 96312368.

RA ALT-MORBE J., STRIKER J.L., FUQUA C., LI P.L., FARRAND S.K.,

RA WINANS S.C.;

RL J. BACTERIOL. 178:4248-4257(1996).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC EMBL; U43675; G2749901; -.

DR CONJUGATION; PLASMD; TRANSMEMBRANE.

KW TRANSMEM 25 45 POTENTIAL.

FT TRANSMEM 307 327 POTENTIAL.

FT TRANSMEM 347 367 POTENTIAL.

SQ SEQUENCE 433 AA; 47258 MW; E3C35AE6 CRC32;

Query Match 70.1%; Score 47; DB 1; Length 433;

Best Local Similarity 70.0%; Pred. No. 1.72e+00;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 79 DGIIGPOQ 88  
|||||  
QY 1 DGTPGEPOQ 10

RESULT 4

ID TIM\_DRONE STANDARD; PRT; 1389 AA.

AC P49021;

DT 01-FEB-1996 (REL. 33, CREATED)

DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)

DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)

DE TIMELESS PROTEIN.

GN TIM.

OS DROSOPHILA MELANOGASTER (FRUIT FLY).

OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 96055118.

RA MYERS M.P., WAGER-SMITH K., WESLEY C.S., YOUNG M.W., SEHGAL A.;

RL SCIENCE 270:805-808(1995).

RN [2]

RP INTERACTION WITH PER.

RX MEDLINE; 96055120.

RA GERAKIS N., SAEZ L., DELAHAYE-BROWN A.M., MYERS M.P., SEHGAL A.,

RA YOUNG M.W., WEITZ C.J.;

RL SCIENCE 270:811-815(1995).

CC -1- FUNCTION: REQUIRED FOR THE PRODUCTION OF CIRCADIAN RHYTHMS.

CC INTERACTS WITH PERIOD (PER). MAY BE REQUIRED AT A SPECIFIC

CC TIME OF DAY TO ALLOW ACCUMULATION AND NUCLEAR LOCALIZATION OF

CC THE PER PROTEIN.

DR EMBL; U37018; G1050970; -.

DR FLYBASE; FBgn0014396; tim.

KW BIOLOGICAL RHYTHMS.

FT DOMAIN 351 380 ASP/GLU-RICH (ACIDIC).

FT DOMAIN 340 353 ARG/LYS-RICH (BASIC).

SQ SEQUENCE 1389 AA; 155665 MW; IDB78941 CRC32;

Query Match 70.1%; Score 47; DB 1; Length 1389;

Best Local Similarity 60.0%; Pred. No. 1.72e+00;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 510 DGPGKPOHQ 519  
||| |  
QY 1 DGTPGEPOQ 10

RESULT 5

ID ERY2\_SACER STANDARD; PRT; 3567 AA.

AC Q03132; Q54096;

DT 01-OCT-1993 (REL. 27, CREATED)

DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE ERYTHRONOLIDE SYNTHASE, MODULES 3 AND 4 (EC 2.3.1.94) (ORF 2) (6-

DE DEOXYERYTHRONOLIDE B SYNTHASE II) (DEBS 2).

GN ERYA.

OS SACCHAROPOLYSPORA ERYTHRAEA (STREPTOMYCES ERYTHRAEUS).

OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; NOCARDIOFORM.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 91220065.

RA DONADIO S., STAVER M.J., MCALPINE J.B., SWANSON S.J., KATZ L.;

RL SCIENCE 252:675-679(1991).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-NRRL 2338;

RX MEDLINE; 92155230.

RA BEVITT D.J., CORTES J., HAYDOCK S.F., LEADLAY P.F.;

RL EUR. J. BIOCHEM. 204:39-49(1992).

CC -1- CATALYTIC ACTIVITY: 6 METHYLMALONYL-COA + PROPIONYL-COA - 7 COA

CC + 6-DEOXYERYTHRONOLIDE B.

CC -1- COFACTOR: NADP.

CC -1- COFACTOR: CONTAINS TWO COVALENTLY BOUND PHOSPHOPANTHEINES.

CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN

CC BIOSYNTHESIS.

CC -1- IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH ENCODING FOR A

CC FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3 ORFS CODES

FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5, AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6, RESPECTIVELY.

-1- BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT), BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR), DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION OF THE FULL-LENGTH CHAIN.

-1- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).

EMBL: M63677; G152694; -.

DR EMBL: X62569; G581651; -.

DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 2.

DR PROSITE: PS00606; B-KETOACYL SYNTHASE; 2.

DR PROSITE: PS50075; ACP\_DOMAIN; 2.

KW TRANSFERASE: ACYLTRANSFERASE; ANTIOTIC BIOSYNTHESIS; NADP;

KW PHOSPHOPANTHETHEINE; MULTIFUNCTIONAL ENZYME.

FT DOMAIN 1 1484 MODULE 3.

FT DOMAIN 1485 3567 MODULE 4.

FT DOMAIN 27 488 BETA-KETOACYL SYNTHASE (KS).

FT DOMAIN 559 884 ACYLTRANSFERASE (AT).

FT DOMAIN 1130 1301 BETA-KETOACYL REDUCTASE (KR) (POSSIBLY NON-FUNCTIONAL).

FT DOMAIN 1397 1467 ACYL CARRIER (ACP).

FT DOMAIN 1485 1943 BETA-KETOACYL SYNTHASE (KS).

FT DOMAIN 2013 2336 ACYLTRANSFERASE (AT).

FT DOMAIN 2383 3066 DEHYDRATASE/ENOYLREDUCTASE (DH/ER).

FT DOMAIN 3139 3322 BETA-KETOACYL REDUCTASE (KR).

FT DOMAIN 3415 3485 ACYL CARRIER (ACP).

FT ACT\_SITE 202 202 THIOESTER BOND.

FT ACT\_SITE 651 651 ACYL-ENZYME INTERMEDIATE.

FT BINDING 1430 1430 PHOSPHOPANTHETHEINE (BY SIMILARITY).

FT ACT\_SITE 1661 1661 THIOESTER BOND.

FT ACT\_SITE 2115 2115 ACYL-ENZYME INTERMEDIATE.

FT NP\_BIND 2961 2978 NADP (ER).

FT NP\_BIND 3142 3157 NADP (KR).

FT BINDING 3448 3448 PHOSPHOPANTHETHEINE (BY SIMILARITY).

FT CONFLICT 438 438 R -> A (IN REF. 2).

FT CONFLICT 480 480 T -> S (IN REF. 2).

FT CONFLICT 1241 1241 L -> F (IN REF. 2).

FT CONFLICT 2664 2664 G -> V (IN REF. 2).

SQ SEQUENCE 3567 AA; 374413 MW; A2F5EA2C CRC32;

Query Match 70.1%; Score 47; DB 1; Length 3567;  
Best Local Similarity 87.5%; Pred. No. 1.72e+00;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1376 GPTGEPAQ 1383  
QY 2 GPTGEPAQ 9  
| | | | | | | |

RESULT 6  
ID SRY\_MUSSI STANDARD; PRT; 311 AA.  
AC Q62563;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DE SEX-DETERMINING REGION Y PROTEIN (TESTIS-DETERMINING FACTOR).  
GN SRY OR TDY.  
OS MUS SPICILEGUS (STEPPE MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98043417.  
RA ALBRECHT K.H., EICHER E.M.;  
RL GENETICS 147:1267-1277(1997).  
RN [2]  
RP SEQUENCE OF 1-143 FROM N.A.  
RX MEDLINE; 93361118.  
RA TUCKER P.K., LUNDRIGAN B.L.;  
RL NATURE 364:715-717(1993).  
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR WHICH REGULATES A GENETIC SWITCH IN MALE DEVELOPMENT. IT IS RESPONSIBLE FOR INITIATING MALE SEX DETERMINATION. SRY HMG BOX RECOGNIZES DNA BY PARTIAL INTERCALATION IN THE MINOR GROOVE.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DOMAIN: THE GLN- AND HIS-RICH DOMAIN MAY MEDIATE PROTEIN-PROTEIN INTERACTIONS.  
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.  
DR EMBL; U70659; G2623383; -.  
DR EMBL; AF009521; G2271483; -.  
DR EMBL; L29544; G496150; -.  
DR MGD; MGI:98860; TDY.  
KW DNA-BINDING; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW SEXUAL DIFFERENTIATION; REPEAT.  
FT DNA\_BIND 5 73 HMG BOX.  
FT VARIANT 274 274 K -> R.  
SQ SEQUENCE 355 AA; 44260 MW; 09CSACAE CRC32;

Query Match 68.7%; Score 46; DB 1; Length 355;  
Best Local Similarity 70.0%; Pred. No. 2.90e+00;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 137 DIPTGHPQQ 146  
QY 1 DGPTGEPAQ 10  
| | | | | | | |

RESULT 7  
ID SRY\_MUSSI STANDARD; PRT; 355 AA.  
AC Q62563;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE SEX-DETERMINING REGION Y PROTEIN (TESTIS-DETERMINING FACTOR).  
GN SRY.  
OS MUS SPRETUS (WESTERN WILD MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98043417.  
RA ALBRECHT K.H., EICHER E.M.;  
RL GENETICS 147:1267-1277(1997).  
RN [2]  
RP SEQUENCE OF 1-143 FROM N.A.  
RX MEDLINE; 93361118.  
RA TUCKER P.K., LUNDRIGAN B.L.;  
RL NATURE 364:715-717(1993).  
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR WHICH REGULATES A GENETIC SWITCH IN MALE DEVELOPMENT. IT IS RESPONSIBLE FOR INITIATING MALE SEX DETERMINATION. SRY HMG BOX RECOGNIZES DNA BY PARTIAL INTERCALATION IN THE MINOR GROOVE.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DOMAIN: THE GLN- AND HIS-RICH DOMAIN MAY MEDIATE PROTEIN-PROTEIN INTERACTIONS.  
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.  
DR EMBL; U70659; G2623383; -.  
DR EMBL; AF009521; G2271483; -.  
DR EMBL; L29544; G496150; -.  
DR MGD; MGI:98860; TDY.  
KW DNA-BINDING; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR;  
KW SEXUAL DIFFERENTIATION; REPEAT.  
FT DNA\_BIND 5 73 HMG BOX.  
FT VARIANT 274 274 K -> R.  
SQ SEQUENCE 311 AA; 38469 MW; 035872C0 CRC32;

Query Match 68.7%; Score 46; DB 1; Length 311;  
Best Local Similarity 70.0%; Pred. No. 2.90e+00;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 137 DIPTGHPQQ 146  
QY 1 DGPTGEPAQ 10  
| | | | | | | |



```
QY 1 DGPTGPEQQE 10

RESULT 8
ID CAL1_RABIT STANDARD; PRT; 53 AA.
AC P02456;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE COLLAGEN ALPHA 1(I) CHAIN (FRAGMENT).
GN COL1A1.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; LAGOMORPHA.
RN [1]
RP SEQUENCE.
RX MEDLINE; 70252720.
RA BORNSTEIN P., NESSE R.;
RL ARCH. BIOCHEM. BIOPHYS. 138:443-450(1970).
CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC (FIBRILLAR FORMING COLLAGEN).
CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC HYDROXYAPATITE.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
DR PIR; A02856; CGBR15.
DR HSP; P12108; IBBF.
DR PROSITE; PS01208; VMFC; PARTIAL.
KW EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;
KW COLLAGEN.
FT MOD_RES 7 7 CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT INVOLVED IN CROSS-LINKING.
FT MOD_RES 26 26 HYDROXYLATION (PROBABLE).
FT MOD_RES 29 29 HYDROXYLATION (PROBABLE).
FT MOD_RES 32 32 HYDROXYLATION (PROBABLE).
FT MOD_RES 41 41 HYDROXYLATION (PROBABLE).
FT MOD_RES 44 44 HYDROXYLATION (PROBABLE).
FT MOD_RES 47 47 HYDROXYLATION (PROBABLE).
FT MOD_RES 53 53 HYDROXYLATION (PROBABLE).
FT NON_TER 53
SQ SEQUENCE 53 AA; 4987 MW; B20D776E CRC32;

Query Match 67.2%; Score 45; DB 1; Length 53;
Best Local Similarity 44.4%; Pred. No. 4.85e+00;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 38 ZGPPGZPGZ 46
QY 1 DGPTGPEQQ 9

RESULT 9
ID YCBC_ECOLI STANDARD; PRT; 259 AA.
AC P36565; P75846;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 28.7 KD PROTEIN IN KDSB-MUKF INTERGENIC REGION.
GN YCBC.
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [2]
RN SEQUENCE OF 1-170 FROM N.A.
RC STRAIN-K12 / W3110;
RX MEDLINE; 94232180.
RA FENG J., YAMANAKA K., NIKI H., OGURA T., HIRAGA S.;

QY 1 DGPTGPEQQE 10

RESULT 11
ID CAL2_HUMAN STANDARD; PRT; 1418 AA.
AC P02458;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROCOLLAGEN ALPHA 1(II) CHAIN PRECURSOR (CONTAINS: CHONDROCALCIN).
GN COL2A1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90067946.
RA SU M.W., LEE B., RAMIREZ F., MACHADO M., HORTON W.;
RL NUCLEIC ACIDS RES. 17:9473-9473(1989).
RN [2]
RN SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE; 87031574.
RA NUNEZ A.M., KOHNO K., MARTIN G.R., YAMADA Y.;
RL GENE 44:11-16(1986).
RN [3]
```

RP SEQUENCE OF 432-1145 FROM N.A.  
RA RAMIREZ F.;  
RL SUBMITTED (DEC-1988) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [4]  
RP SEQUENCE OF 963-1418 FROM N.A.  
RX MEDLINE; 85190334.  
RA CHEAH K.S.E., STOKER N.G., GRIFFIN J.R., GROSVELD F.G., SOLOMON E.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 82:2555-2559(1985).  
RN [5]  
RP SEQUENCE OF 1120-1398 FROM N.A.  
RX MEDLINE; 85306861.  
RA ELIMA K., MAEKELAE J.K., VUORIO T., KAUPPINEN S., KNOWLES J.,  
RA VUORIO E.;  
RL BIOCHEM. J. 229:183-188(1985).  
RN [6]  
RP SEQUENCE OF 1106-1418 FROM N.A.  
RX MEDLINE; 88067771.  
RA ELIMA K., VUORIO T., VUORIO E.;  
RL NUCLEIC ACIDS RES. 15:9499-9504(1987).  
RN [7]  
RP SEQUENCE OF 1227-1289 FROM N.A.  
RX MEDLINE; 86104139.  
RA NUNEZ A.M., FRANCOMANO C., YOUNG M.F., MARTIN G.R., YAMADA Y.;  
RL BIOCHEMISTRY 24:6343-6348(1985).  
RN [8]  
RP SEQUENCE OF 1176-1226 FROM N.A.  
RX MEDLINE; 84118798.  
RA STROM C.M., UPHOLT W.B.;  
RL NUCLEIC ACIDS RES. 12:1025-1038(1984).  
RN [9]  
RP SEQUENCE OF 35-167 FROM N.A.  
RX MEDLINE; 89233138.  
RA SU M.W., BENSON-CHANDA V., VISSING H., RAMIREZ F.;  
RL GENOMICS 4:438-441(1989).  
RN [10]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 91184577.  
RA KUIVANEMI H., TROMP G., PROCKOP D.J.;  
RL FASEB J. 5:2052-2060(1991).  
RN [11]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 97259599.  
RA KUIVANEMI H., TROMP G., PROCKOP D.J.;  
RL HUM. MUTAT. 9:300-315(1997).  
RN [12]  
RP VARIANT SER-1074.  
RX MEDLINE; 90036909.  
RA VISSING H., D'ALESSIO M., LEE B., RAMIREZ F., GODFREY M.,  
RA HOLLISTER D.W.;  
RL J. BIOL. CHEM. 264:18265-18267(1989).  
RN [13]  
RP VARIANT SEDC GLY-1095--TYR-1330 DEL.  
RX MEDLINE; 89266907.  
RA LEE B., VISSING H., RAMIREZ F., ROGERS D., RIMOIN D.;  
RL SCIENCE 244:978-980(1989).  
RN [14]  
RP VARIANT OSTEOARTHRITIS CYS-650.  
RX MEDLINE; 90370826.  
RA ALA-KOKKO L., BALDWIN C.T., MOSKOWITZ R.W., PROCKOP D.J.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 87:6565-6568(1990).  
RN [15]  
RP VARIANT OI-IV VAL-717.  
RX MEDLINE; 91291136.  
RA BATEMAN J.F., HANNAGAN M., CHAN D., COLE W.G.;  
RL BIOCHEM. J. 276:765-770(1991).  
RN [16]  
RP VARIANT OSTEOARTHRITIS CYS-650.  
RX MEDLINE; 91086471.  
RA EYRE D.R., WEIS M.A., MOSKOWITZ R.W.;  
RL J. CLIN. INVEST. 87:357-361(1991).  
RN [17]  
RP VARIANT HYPOCHONDROGENESIS GLU-984.  
RX MEDLINE; 93054548.

RA BOGAERT R., TILLER G.E., WIES M.A., GRUBER H.E., RIMOIN D.L.,  
RA COHN D.H., EYRE D.R.;  
RL J. BIOL. CHEM. 267:22522-22526(1992).  
RN [18]  
RP VARIANT HYPOCHONDROGENESIS SER-705.  
RX MEDLINE; 92262484.  
RA HORTON W.A., MACHADO M.A., ELLARD J., CAMPBELL D., BARTLEY J.,  
RA RAMIREZ F., VITALE E., LEE B.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 89:4583-4587(1992).  
RN [19]  
RP VARIANT WS-II ASP-198.  
RX MEDLINE; 93304428.  
RA KOKKO J., RITVANEMI P., HAATAJA L., KAARIAINEN H., KIVIRIKKO K.I.,  
RA PROCKOP D.J., ALA-KOKKO L.;  
RL AM. J. HUM. GENET. 53:55-61(1993).  
RN [20]  
RP VARIANT SEMD CYS-940.  
RA TILLER G.E., WEIS M.A., LACHMAN R.S., COHN D.H., RIMOIN D.L.,  
RA EYRE D.R.;  
RL AM. J. HUM. GENET. 53:209-209(1993).  
RN [21]  
RP VARIANT OSTEOARTHRITIS CYS-650.  
RX MEDLINE; 93282819.  
RA HODDERBAUM D., MALEMUD C.J., MOSKOWITZ R.W., HAQOI T.M.;  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 192:1169-1174(1993).  
RN [22]  
RP VARIANT SEMD ARG-285.  
RX MEDLINE; 93252400.  
RA VIRKULA M., RITVANEMI P., VUORIO A.F., KAITILA I., ALA-KOKKO L.,  
RA PELTONEN L.;  
RL GENOMICS 16:282-285(1993).  
RN [23]  
RP VARIANT SEDC CYS-206.  
RX MEDLINE; 94063862.  
RA WILLIAMS C.J., CONSIDINE E.L., KNOWLTON R.G., REGINATO A., NEUMANN G.,  
RA HARRISON D., BUXTON P., JIMENEZ S., PROCKOP D.J.;  
RL HUM. GENET. 92:499-505(1993).  
RN [24]  
RP VARIANT SEDC CYS-920.  
RX MEDLINE; 93315508.  
RA CHAN D., TAYLOR T.K.F., COLE W.G.;  
RL J. BIOL. CHEM. 268:15238-15245(1993).  
RN [25]  
RP VARIANT SEDC SER-1128.  
RX MEDLINE; 93140139.  
RA COLE W.G., HALL R.K., ROGERS J.G.;  
RL J. MED. GENET. 30:27-35(1993).  
RN [26]  
RP VARIANT AOM ALA-233--LYS-239 DEL.  
RX MEDLINE; 95067975.  
RA BOGAERT R., WILKIN D., WILCOX W.R., LACHMAN R., RIMOIN D., COHN D.H.,  
RA EYRE D.R.;  
RL AM. J. HUM. GENET. 55:1128-1136(1994).  
RN [27]  
RP VARIANT KS ASP-234.  
RX MEDLINE; 95179117.  
RA WILKIN D.J., BOGAERT R., LACHMAN R.S., RIMOIN D.L., EYRES D.R.,  
RA COHN D.H.;  
RL HUM. MOL. GENET. 3:1999-2003(1994).  
RN [28]  
RP VARIANT OI-III VAL-717.  
RX MEDLINE; 95187161.  
RA FORLINO A., ZOLEZZI F., VALLI M., PIGNATTI P.F., CETTA G.,  
RA BRUNELLI P.C., MOTTESS M.;  
RL HUM. MOL. GENET. 3:2201-2206(1994).  
RN [29]  
RP VARIANT SEDC SER-378.  
RX MEDLINE; 94290498.  
RA RITVANEMI P., SOKOLOV B.P., WILLIAMS C.J., CONSIDINE W., YURGENEV L.,  
RA MEERSON E.M., ALA-KOKKO L., PROCKOP D.J.;  
RL HUM. MUTAT. 3:261-267(1994).  
RN [30]  
RP VARIANTS CYS-206; CYS-650; ARG-822; SER-1107 AND ARG-1119.

RX MEDLINE; 95276749.  
 RA WILLIAMS C.J., ROCK M., CONSIDINE E., MCCARRON S., GOW P., LADDA R.,  
 RA MCCLAIN D., MICHELS V.M., MURPHY W., PROCKOP D.J., GANGULY A.;  
 RL HUM. MOL. GENET. 4:309-312(1995).  
 RN [31]  
 RP VARIANT HYPOCHONDROGENESIS CYS-1044.  
 RX MEDLINE; 96298263.  
 RA MUNDLOS S., CHAN D., MCGILL J., BATEMAN J.F.;  
 RL AM. J. MED. GENET. 63:129-136(1996).  
 RN [32]  
 RP VARIANT KS PRO-1138--GLY-1143 DEL.  
 RX MEDLINE; 97016530.  
 RA WINTERPACHT A., SUPERTI-FURGA A., SCHWARZE U., STOESE H.,  
 RA STEINMANN B., SPRANGER J., ZABEL B.;  
 RL J. MED. GENET. 33:649-654(1996).  
 CC -!- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.  
 CC -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.  
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -!- DISEASE: DEFECTS IN COL2A1 ARE THE CAUSE OF A VARIETY OF  
 CC CHONDRODYSPLASIA INCLUDING HYPOCHONDROGENESIS AND OSTEOARTHRITIS.  
 CC -!- DISEASE: DEFECTS IN COL2A1 ARE THE CAUSE OF SPONDYLOEPHYSAL  
 CC DYSPLASIA, CONGENITAL TYPE (SDC); A DISORDER CHARACTERIZED BY  
 CC DISPROPORTIONATE SHORT STATURE AND PLEIOTROPIC INVOLVEMENT OF THE  
 CC SKELETAL AND OCULAR SYSTEMS.  
 CC -!- DISEASE: DEFECTS IN COL2A1 ARE THE CAUSE OF TYPE II WAGNER  
 CC SYNDROME (WS-II); A DISEASE CHARACTERIZED BY EARLY-ONSET  
 CC CATARACTS, LATTICE DEGENERATION OF THE RETINA, AND RETINAL  
 CC DETACHMENT WITHOUT INVOLVEMENT OF MONOCULAR TISSUES.  
 CC -!- DISEASE: DEFECTS IN COL2A1 ARE THE CAUSE OF TYPE II  
 CC ACHONDROGENESIS-HYPOCHONDROGENESIS; A DISEASE CHARACTERIZED BY THE  
 CC ABSENCE OF OSSIFICATION IN THE VERTEBRAL COLUMN, SACRUM AND PUBIC  
 CC BONES.  
 CC  
 Note: remainder of annotations omitted.  
 Query Match 67.2%; Score 45; DB 1; Length 1418;  
 Best Local Similarity 66.7%; Pred. No. 4.85e+00;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 930 GPSGPGQQ 938  
 QY 2 GTGEPQQE 10  
 |||||  
 RESULT 12  
 ID MAM\_DROME STANDARD; PRT; 1596 AA.  
 AC P21519;  
 DT 01-MAY-1991 (REL. 18, CREATED)  
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)  
 DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)  
 DE NEUROGENIC PROTEIN MASTERMIND.  
 GN MAM.  
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CANTON-S;  
 RX MEDLINE; 91065516.  
 RA SMOLLER D., FRIEDEL C., SCHMID A., BETTLER D., LAM L.,  
 RA YEDVOBNICK B.;  
 RL GENES DEV. 4:1688-1700(1990).  
 CC -!- FUNCTION: MAY HAVE A REGULATORY FUNCTION POSSIBLY IN ASSOCIATION  
 CC WITH THE N GENE PRODUCT.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- DEVELOPMENTAL STAGE: DURING EARLY NEUROGENESIS MAM PRODUCTS ARE  
 CC UBQUITOUSLY LOCATED. DURING LATER STAGES THEY ACCUMULATE IN THE  
 CC CENTRAL NERVOUS SYSTEM.  
 CC -!- THE PROTEIN HAS MANY AA HOMOPOLYMERIC DOMAINS: 21 POLY-GLN RUNS  
 CC (FROM 5 TO 16 AA IN LENGTH), 4 POLY-GLY (6 TO 10 AA), 3 POLY-ASN  
 CC (3 X 5 AA), 1 POLY-ALA (10 AA) AND 1 POLY-THR (5 AA) RUNS.  
 CC -!- SIMILARITY: TO OTHER NUCLEAR PROTEINS OF DROSOPHILA, TO CERTAIN  
 CC YEAST AND MAMMALIAN REGULATORY PROTEINS.

DR EMBL; X54251; G8204; -  
 DR PIR; A33106; A33106.  
 DR PIR; A36391; A36391.  
 DR FLYBASE; FBgn0002643; mam.  
 KW NEUROGENESIS; NUCLEAR PROTEIN; REPEAT.  
 FT DOMAIN 20 84  
 FT DOMAIN 127 190  
 FT DOMAIN 196 219  
 FT DOMAIN 259 304  
 FT DOMAIN 355 388  
 FT DOMAIN 392 406  
 FT DOMAIN 407 440  
 FT DOMAIN 651 671  
 FT DOMAIN 700 714  
 FT DOMAIN 759 816  
 FT DOMAIN 987 996  
 FT DOMAIN 1060 1079  
 FT DOMAIN 1092 1107  
 FT DOMAIN 1237 1252  
 FT DOMAIN 1492 1496  
 FT DOMAIN 1559 1592  
 SQ SEQUENCE 1596 AA; 167717 MW; BEB0500 CRC32;  
 Query Match 67.2%; Score 45; DB 1; Length 1596;  
 Best Local Similarity 66.7%; Pred. No. 4.85e+00;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Db 1317 GPMGGPQQ 1325  
 QY 2 GTGEPQQE 10  
 |||||  
 RESULT 13  
 ID PRP2\_MOUSE STANDARD; PRT; 261 AA.  
 AC P05142;  
 DT 13-AUG-1987 (REL. 05, CREATED)  
 DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
 DE PROLINE-RICH PROTEIN MP-2 PRECURSOR.  
 GN PRP.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC ANN D.K., CARLSON D.M.;  
 RX MEDLINE; 86059475.  
 RL J. BIOL. CHEM. 260:15863-15872(1985).  
 DR EMBL; M12099; G200547; -  
 DR HSSP; P19999; ICLG.  
 KW REPEAT; SALIVA; SIGNAL.  
 FT SIGNAL 1 15  
 FT CHAIN 16 261  
 FT SEQUENCE 261 AA; 26034 MW; 9D830DAF CRC32;  
 Query Match 65.7%; Score 44; DB 1; Length 261;  
 Best Local Similarity 75.0%; Pred. No. 8.07e+00;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Db 193 PTGPGPQQ 200  
 QY 3 PTGEPQQE 10  
 |||||  
 RESULT 14  
 ID ALD\_HUMAN STANDARD; PRT; 745 AA.  
 AC P33897;  
 DT 01-FEB-1994 (REL. 28, CREATED)  
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE ADRENOLEUKODYSTROPHY PROTEIN (ALDP).  
 GN ALD.  
 OS HOMO SAPIENS (HUMAN).

CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 93180910.  
 RA MOSSER J., DOUAR A.-M., SARDE C.-O., KIOSCHIS P., FEIL R., MOSER H.,  
 RA POUSTKA A.-M., MANDEL J.-L., AUBOURG P.;  
 RL NATURE 361:726-730(1993).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA PLATZER M., BAUER D., BRENNER V., DRESCHER B., NYAKATURA G.,  
 RA REICHWALD K., SANDOVAL N., COY J., KIOSCHIS P., KORN B.,  
 RA POUSTKA A.-M., ROSENTHAL A.;  
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 [3]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE: 97338663.  
 RA DODD A., ROWLAND S.A., HAWKES S.L.J., KENNEDY M.A., LOVE D.R.;  
 RL HUM. MUTAT. 9:500-511(1997).  
 [4]  
 RP VARIANT X-ALD LYS-291.  
 RX MEDLINE: 94108454.  
 RA CARTIER N., SARDE C.-O., DOUAR A.-M., MOSSER J., MANDEL J.-L.,  
 RA AUBOURG P.;  
 RL HUM. MOL. GENET. 2:1949-1951(1993).  
 [5]  
 RP REVIEW.  
 RX MEDLINE: 93283453.  
 RA AUBOURG P., MOSSER J., DOUAR A.-M., SARDE C.-O., LOPEZ J.,  
 RA MANDEL J.-L.;  
 RL BIOCHIMIE 75:293-302(1993).  
 [6]  
 RP VARIANTS X-ALD SER-148; ASP-174; ARG-266; GLN-401; TRP-418 & PHE-515.  
 RX MEDLINE: 95152524.  
 RA FUCHS S., SARDE C.-O., WEDEMANN H., SCHWINGER E., MANDEL J.-L.,  
 RA GAL A.;  
 RL HUM. MOL. GENET. 3:1903-1905(1994).  
 [7]  
 RP VARIANTS X-ALD TRP-518; LEU-606; CYS-617 AND HIS-617.  
 RX MEDLINE: 94314951.  
 RA FANEN P., GUIDOUX S., SARDE C.-O., MANDEL J.-L., GOOSSENS M.,  
 RA AUBOURG P.;  
 RL J. CLIN. INVEST. 94:516-520(1994).  
 [8]  
 RP VARIANTS X-ALD.  
 RX MEDLINE: 95126139.  
 RA LIGTENBERG M.J.L., KEMP S., SARDE C.-O., VAN GEEL B.M., KLEIJER W.J.,  
 RA BARTH P.G., MANDEL J.-L., VAN OOST B.A., BOLHUIS P.A.;  
 RL AM. J. HUM. GENET. 56:44-50(1995).  
 [9]  
 RP VARIANTS X-ALD HIS-104; GLU-178; LEU-560 AND GLY-528 DEL.  
 RX MEDLINE: 95233433.  
 RA BRAUN A., AMBACH H., KAMMERER S., ROLINSKI B., STOECKLER S., RABL W.,  
 RA GAERTNER J., ZIERZ S., ROSCHER A.A.;  
 RL AM. J. HUM. GENET. 56:854-861(1995).  
 [10]  
 RP VARIANTS X-ALD.  
 RX MEDLINE: 96047143.  
 RA KOK F., NEUMANN S., SARDE C.-O., ZHENG S., WU K.-H., WEI H.-M.,  
 RA BERGIN J., WATKINS P.A., GOULD S., SACK G., MOSER H., MANDEL J.-L.,  
 RA SMITH K.D.;  
 RL HUM. MUTAT. 6:104-115(1995).  
 [11]  
 RP VARIANTS X-ALD.  
 RX MEDLINE: 96213748.  
 RA FEIGENBAUM V., LOMBARD-PLATET G., GUIDOUX S., SARDE C.-O.,  
 RA MANDEL J.-L., AUBOURG P.;  
 RL AM. J. HUM. GENET. 58:1135-1144(1996).  
 [12]  
 RP VARIANTS X-ALD.  
 RX MEDLINE: 96163493.  
 RA KRAEMANN E.W., MEIER V., KORENKE G.C., HUNNEMAN D.H., HANEFELD F.;  
 RL HUM. GENET. 97:194-197(1996).

[13]  
 RN VARIANT AMN ARG-679.  
 RP KORENKE G.C., KRAEMANN E., MEIER V., BEUCHE W., HUNNEMAN D.H.,  
 RA HANEFELD F.;  
 RL HUM. MUTAT. SUPPL. 1:S204-S206(1998).  
 CC -I- FUNCTION: PROBABLE TRANSPORTER. COULD BE INVOLVED IN THE  
 CC IMPORT OF VLCFA-COA SYNTHETASE INTO THE PEROXISOMAL MEMBRANE.  
 CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL  
 CC (PROBABLE).  
 CC -I- DISEASE: DEFECTS IN ALD ARE THE CAUSE OF RECESSIVE X-LINKED  
 CC ADRENOLEUKODYSTROPHY (X-ALD), A RARE PEROXISOMAL METABOLIC  
 CC DISORDER THAT OCCURS IN BOYS AND IS CHARACTERIZED BY PROGRESSIVE  
 CC MULTIFOCAL DEMYELINATION OF THE CENTRAL NERVOUS SYSTEM AND BY  
 CC ADRENOCORTICAL INSUFFICIENCY. IT PRODUCES MENTAL DETERIORATION,  
 CC CORTICOSPINAL TRACT DYSFUNCTION, AND CORTICAL BLINDNESS. THERE IS  
 CC LABORATORY EVIDENCE OF ADRENAL CORTICAL DYSFUNCTION. DEATH  
 CC INVARIABLY OCCURS IN 1 TO 5 YEARS. DIFFERENT CLINICAL  
 CC MANIFESTATIONS EXIST LIKE: CEREBRAL CHILDHOOD ALD (CALD), ADULT  
 CC CEREBRAL ALD (ACALD), ADRENOMYELONEUROPATHY (AMN) AND "ADDISON  
 CC DISEASE ONLY" (ADO) PHENOTYPE.  
 CC -I- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
 CC (ABC TRANSPORTERS). BELONGS TO THE MDR SUBFAMILY.  
 DR EMBL: Z21876; G38591.  
 DR EMBL: Z31348; G806557.  
 DR EMBL: Z31006; G806557; JOINED.  
 DR EMBL: Z31007; G806557; JOINED.  
 DR EMBL: Z31008; G806557; JOINED.  
 DR EMBL: Z31009; G806557; JOINED.  
 DR EMBL: Z31010; G806557; JOINED.  
 DR EMBL: U52111; G1302652.  
 DR PIR: S30059; S30059.  
 DR HSSP: P02392; ICTF.  
 DR MIM: 300100.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 KW ATP-BINDING; GLYCOPROTEIN; TRANSMEMBRANE; TRANSPORT; PEROXISOME;  
 KW DISEASE MUTATION; POLYMORPHISM.  
 FT TRANSMEM 92 112 POTENTIAL.  
 FT TRANSMEM 131 151 POTENTIAL.  
 FT TRANSMEM 238 258 POTENTIAL.  
 FT TRANSMEM 333 353 POTENTIAL.  
 FT TRANSMEM 473 493 POTENTIAL.  
 FT NP\_BIND 507 514 ATP (BY SIMILARITY).  
 FT VARIANT 98 98 S -> L (IN X-ALD; CALD TYPE).  
 FT VARIANT 104 104 R -> C (IN X-ALD).  
 FT VARIANT 104 104 R -> H (IN X-ALD; ADO-TYPE).  
 FT VARIANT 105 105 T -> I (IN X-ALD; ADO-TYPE).  
 FT VARIANT 107 107 L -> P (IN X-ALD; ALD/AMN/ADO-TYPES AND  
 FT VARIANT 107 107 ASSYMPTOMATIC).  
 FT VARIANT 108 108 S -> W (IN X-ALD; CALD AND AMN-TYPES).  
 FT VARIANT 116 116 G -> R (IN X-ALD; CALD-TYPE).  
 FT VARIANT 123 123 A -> V.  
 FT VARIANT 138 141 MISSING (IN X-ALD; ALD-TYPE).  
 FT VARIANT 141 141 A -> T (IN X-ALD).  
 FT VARIANT 148 148 N -> S (IN X-ALD; ADO-TYPE).  
 FT VARIANT 149 149 S -> N (IN X-ALD).  
 FT VARIANT 152 152 R -> C (IN X-ALD; ADO-TYPE).  
 FT VARIANT 152 152 R -> P (IN X-ALD).  
 FT VARIANT 163 163 R -> H (IN X-ALD).  
 FT VARIANT 174 174 Y -> D (IN X-ALD; ALD-TYPE).  
 FT VARIANT 174 174 Y -> S (IN X-ALD; CALD-TYPE).  
 FT VARIANT 178 178 Q -> E (IN X-ALD; AMN-TYPE).  
 FT VARIANT 181 181 Y -> C (IN X-ALD; ALMD-TYPE).  
 FT VARIANT 182 182 R -> P (IN X-ALD).  
 FT VARIANT 194 194 D -> H (IN X-ALD).  
 FT VARIANT 200 200 D -> V (IN X-ALD; CALD-TYPE).  
 FT VARIANT 211 211 L -> P (IN X-ALD).  
 FT VARIANT 220 220 L -> P (IN X-ALD).  
 FT VARIANT 221 221 D -> G (IN X-ALD; CALD AND AMN-TYPES).  
 FT VARIANT 254 254 T -> M (IN X-ALD; AMN-TYPE).  
 FT VARIANT 254 254 T -> P (IN X-ALD; AMN-TYPE).  
 FT VARIANT 263 263 P -> L (IN X-ALD; CALD, AMN AND AD-  
 FT VARIANT 263 263 TYPES).  
 FT VARIANT 266 266 G -> R (IN X-ALD).

FT VARIANT 276 K -> E (IN X-ALD; CALD-TYPE).  
FT VARIANT 277 G -> R (IN X-ALD; AMN-TYPE).  
FT VARIANT 277 G -> GN (IN X-ALD; ADO-TYPE).  
FT VARIANT 277 G -> W (IN X-ALD).  
FT VARIANT 291 E -> D (IN X-ALD; ACALD AND CALD-TYPES).  
FT VARIANT 291 E -> K (IN X-ALD).  
FT VARIANT 291 MISSING (IN X-ALD; ALD-TYPE).  
FT VARIANT 291 A -> T (IN X-ALD; AMN-TYPE).  
FT VARIANT 294 S -> P (IN X-ALD; AMN-TYPE).  
FT VARIANT 342 S -> G (IN X-ALD; AMN-TYPE).  
FT VARIANT 389 R -> H (IN X-ALD).  
FT VARIANT 389 R -> Q (IN X-ALD; ALD AND AMN-TYPES).  
FT VARIANT 401 R -> W (IN X-ALD; AMN-TYPE).  
FT VARIANT 418 P -> R (IN X-ALD; CALD, AMN AND  
FT VARIANT 484 ADO-TYPES).  
FT VARIANT 507 G -> V (IN X-ALD; CALD-TYPES).  
FT VARIANT 512 G -> S (IN X-ALD; CALD AND AS-TYPES).  
FT VARIANT 515 S -> F (IN X-ALD).  
FT VARIANT 518 R -> Q (IN X-ALD; CALD-TYPE).  
FT VARIANT 518 R -> W (IN X-ALD; CALD-TYPE).  
FT VARIANT 522 G -> W (IN X-ALD; AD-TYPE).  
FT VARIANT 528 MISSING (IN X-ALD; CALD-TYPE).  
FT VARIANT 534 P -> L (IN X-ALD; CALD-TYPE).  
FT VARIANT 560 P -> L (IN X-ALD; CALD-TYPE).  
FT VARIANT 560 P -> R (IN X-ALD; AMN AND ALMD-TYPES).  
FT VARIANT 566 M -> K (IN X-ALD).  
FT VARIANT 591 R -> Q (IN X-ALD; AMN-TYPE).  
FT VARIANT 606 S -> L (IN X-ALD).  
FT VARIANT 606 S -> P (IN X-ALD; CALD, AMN AND ALMD-  
FT VARIANT 609 TYPES).  
FT VARIANT 609 E -> G (IN X-ALD).  
FT VARIANT 609 E -> K (IN X-ALD; AMN-TYPE).  
FT VARIANT 617 R -> C (IN X-ALD; ALD-TYPE AND  
FT VARIANT 617 ASYMPTOMATIC).  
FT VARIANT 617 R -> G (IN X-ALD; ADO AND AMN-TYPES WITH  
FT VARIANT 617 CEREBRAL INVOLVEMENT).  
FT VARIANT 617 R -> H (IN X-ALD).  
FT VARIANT 626 A -> T (IN X-ALD; CALD AND AMN-TYPES).  
FT VARIANT 629 D -> H (IN X-ALD).  
FT VARIANT 657 MISSING (IN X-ALD; CALD-TYPE).  
FT VARIANT 660 R -> W (IN X-ALD; CALD, ALMD AND AS-  
FT VARIANT 679 TYPES).  
FT VARIANT 679 W -> R (IN X-ALD; AMN-TYPE).  
FT VARIANT 745 AA; 82908 MW; F37A45D3 CRC32;  
SQ SEQUENCE

Query Match 65.7%; Score 44; DB 1; Length 745;  
Best Local Similarity 60.0%; Pred. No. 8.07e+00;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 QAPAGEPTQE 56  
QY 1 DGPTGEPTQE 10

RESULT 15  
ID RRE1\_HUMAN STANDARD; PRT; 755 AA.  
AC Q92766;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE RAS-RESPONSIVE ELEMENT BINDING PROTEIN 1 (RREB-1).  
GN RREB1.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-THYROID CARCINOMA;  
RX MEDLINE; 96413283.  
RA THIAGALINGAM A.; BAYLIN S.B.; BORGES M.; NELKIN B.D.;  
RL MOL. CELL. BIOL. 16:5335-5345(1996).  
CC -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS SPECIFICALLY TO THE  
DISTAL RAS-RESPONSIVE ELEMENT (RRE) IN THE CALCITONIN GENE

CC PROMOTER AND AUGMENT THE RAS/RAF-MEDIATED TRANSCRIPTIONAL RESPONSE  
CC OF THAT PROMOTER. MAY BE INVOLVED IN RAS/RAF-MEDIATED CELL  
CC DIFFERENTIATION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL; U26914; G1654112; -.  
DR MIM; 602209; -.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2; 4.  
KW TRANSCRIPTION REGULATION; ACTIVATOR; ZINC-FINGER; METAL-BINDING;  
KW DNA-BINDING; NUCLEAR PROTEIN.  
FT ZN\_FING 315 337 C2H2-TYPE.  
FT ZN\_FING 461 483 C2H2-TYPE.  
FT ZN\_FING 580 602 C2H2-TYPE.  
FT ZN\_FING 608 630 C2H2-TYPE.  
SQ SEQUENCE 755 AA; 79865 MW; 50627236 CRC32;

Query Match 65.7%; Score 44; DB 1; Length 755;  
Best Local Similarity 50.0%; Pred. No. 8.07e+00;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 501 EGSPAPEQE 510  
QY 1 DGPTGEPTQE 10

Search completed: Thu May 13 15:37:02 1999  
Job time : 8 secs.

\*\*\*\*\*

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

MPPerch_pp  protein - protein database search, using Smith-Waterman algorithm
Run on:      Thu May 13 15:37:21 1999;  MasPar time 3.98 Seconds
Tabular output not generated.  125.109 Million cell updates/sec

```

```
>US-09-040-485-8
Description: (1-10) from US09040485.pep (1 of 2)
Perfect Score: 67
Sequence: 1 DGPTEPQQE 10
```

Scoring table: PAM 150  
Gap 15

Searched: 165420 seqs, 49795644 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

spiremb16

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 20.459; Variance 23.634; scale 0.866

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query |      | Length | DB     | ID                      | Description | Pred. No. |
|------------|-------|-------|------|--------|--------|-------------------------|-------------|-----------|
|            |       | Match | %    |        |        |                         |             |           |
| 1          | 51    | 76.1  | 205  | 5      | O01718 | 23KDA HEAT SHOCK PROTE  | 3.45e+01    |           |
| 2          | 48    | 71.6  | 414  | 2      | P73855 | HYPOTHETICAL 44.8 KD P  | 1.74e+00    |           |
| 3          | 47    | 70.1  | 357  | 13     | P79765 | NEUROD-LIKE PROTEIN.    | 2.95e+00    |           |
| 4          | 47    | 70.1  | 357  | 13     | P73826 | NEUROD.                 | 2.95e+00    |           |
| 5          | 47    | 70.1  | 676  | 5      | O44431 | CIRCADIAN CLOCK PROTEI  | 2.95e+00    |           |
| 6          | 47    | 70.1  | 1120 | 5      | O20778 | SIMILAR TO TRIPLE HELI  | 2.95e+00    |           |
| 7          | 47    | 70.1  | 1194 | 5      | O44430 | CIRCADIAN CLOCK PROTEI  | 2.95e+00    |           |
| 8          | 47    | 70.1  | 1343 | 5      | O17482 | TIMELESS (TIM).         | 2.95e+00    |           |
| 9          | 47    | 70.1  | 1398 | 5      | O44380 | CIRCADIAN CLOCK PROTEI  | 2.95e+00    |           |
| 10         | 46    | 68.7  | 231  | 11     | O35966 | SEX DETERMINING PROTEI  | 4.96e+00    |           |
| 11         | 46    | 68.7  | 232  | 11     | O35968 | SEX DETERMINING PROTEI  | 4.96e+00    |           |
| 12         | 46    | 68.7  | 233  | 11     | O35860 | SEX DETERMINING PROTEI  | 4.96e+00    |           |
| 13         | 46    | 68.7  | 234  | 11     | O35858 | SEX DETERMINING PROTEI  | 4.96e+00    |           |
| 14         | 46    | 68.7  | 234  | 11     | O35859 | SEX DETERMINING PROTEI  | 4.96e+00    |           |
| 15         | 46    | 68.7  | 238  | 2      | O69004 | FERRISIDEROPHORE RECEP  | 4.96e+00    |           |
| 16         | 45    | 67.2  | 89   | 7      | O19497 | MHC CLASS II BETA 1 DO  | 8.28e+00    |           |
| 17         | 45    | 67.2  | 227  | 7      | O31398 | MHC CLASS II B-L BETA   | 8.28e+00    |           |
| 18         | 45    | 67.2  | 263  | 7      | O31410 | MHC CLASS II B-LBII1-BE | 8.28e+00    |           |
| 19         | 45    | 67.2  | 263  | 13     | O73896 | B LOCUS L BETA CHAIN 2  | 8.28e+00    |           |
| 20         | 45    | 67.2  | 488  | 14     | O37935 | POLYPROTEIN (FRAGMENT)  | 8.28e+00    |           |

## ALIGNMENTS

|        |   |
|--------|---|
| RESULT | 1   |
| ID     | 001718 PRELIMINARY; PRT; 205 AA.                                      |
| AC     | 001718;   |
| DT     | 01-JUL-1997 (TREMBREL. 04, CREATED)                                   |
| DT     | 01-JUL-1997 (TREMBREL. 04, LAST SEQUENCE UPDATE)                      |
| DT     | 01-JUN-1998 (TREMBREL. 06, LAST ANNOTATION UPDATE)                    |
| DE     | 23KDA HEAT SHOCK PROTEIN SCHP23.                                      |
| DE     | SARCOPHAGA CRASSIPALPIS.  |
| OS     | EUKARYOTAE; MITOCHONDRIAL EUKARYOTES; METAZOA; ARTHROPODA; TRACHEATA; |
| OC     | HEXAPODA; INSECTA; PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA;       |
| OC     | OESTRIDEA; SARCOPHAGIDAE; SARCOPHAGA.                                 |
| RN     | [1]   |
| RP     | SEQUENCE FROM N.A.  |
| RA     | VOCUM G.D., JOPLIN K.H., DENLINGER D.L.;                              |
| RL     | SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.                 |
| DR     | EMBL; U96099; G2038737; -   |
| DR     | PFAM; PF00011; HSP20.   |
| KW     | HEAT SHOCK.   |
| SQ     | SEQUENCE 205 AA; 22933 MW; 5164AA349 CRC32:                           |

|    |    |      |         |        |                         |          |
|----|----|------|---------|--------|-------------------------|----------|
| 21 | 44 | 55.7 | 243.14  | Q67788 | 26 KD PROTEIN.          | 1.37e+01 |
| 22 | 44 | 55.7 | 295.5   | Q20927 | F57B7.3.                | 1.37e+01 |
| 23 | 44 | 55.7 | 306.5   | Q20135 | F38A3.1.                | 1.37e+01 |
| 24 | 44 | 65.7 | 310.5   | Q20282 | SIMILAR TO CURTICLE COL | 1.37e+01 |
| 25 | 44 | 65.7 | 319.5   | Q19707 | F2285.3.                | 1.37e+01 |
| 26 | 44 | 65.7 | 389.5   | Q21975 | SIMILAR TO C. ELEGANS   | 1.37e+01 |
| 27 | 44 | 65.7 | 474.5   | Q21975 | KINESIN-RELATED PROTEI  | 1.37e+01 |
| 28 | 44 | 65.7 | 632.4   | Q06887 | CICKO721Q.3 (KINESIN R  | 1.37e+01 |
| 29 | 44 | 65.7 | 934.2   | P73172 | CHEMOTAXIS PROTEIN CHE  | 1.37e+01 |
| 30 | 44 | 65.7 | 1110.13 | Q91255 | NF-180.                 | 1.37e+01 |
| 31 | 44 | 65.7 | 1142.4  | Q14697 | B120.                   | 1.37e+01 |
| 32 | 44 | 65.7 | 1302.4  | Q14525 | KIRA0289 (FRAGMENT).    | 1.37e+01 |
| 33 | 43 | 64.2 | 89.7    | Q19499 | MHC CLASS II BETA 1 DO  | 2.26e+01 |
| 34 | 43 | 64.2 | 89.7    | Q19492 | MHC CLASS II BETA 1 DO  | 2.26e+01 |
| 35 | 43 | 64.2 | 115.11  | Q35230 | KINESIN MOTOR PROTEIN   | 2.26e+01 |
| 36 | 43 | 64.2 | 163.11  | Q54719 | KINESIN-RELATED PROTEI  | 2.26e+01 |
| 37 | 43 | 64.2 | 266.7   | Q31414 | MHC CLASS II BETA CHAI  | 2.26e+01 |
| 38 | 43 | 64.2 | 306.4   | Q16258 | COLLAGEN ALPHA 1(VI) C  | 2.26e+01 |
| 39 | 43 | 64.2 | 313.5   | Q61525 | F17P9.1 PROTEIN.        | 2.26e+01 |
| 40 | 43 | 64.2 | 389.1   | Q34135 | 40 KDA HEAT SHOCK CHAP  | 2.26e+01 |
| 41 | 43 | 64.2 | 394.5   | Q46068 | 1-EVIDENCE-PREDICTED B  | 2.26e+01 |
| 42 | 43 | 64.2 | 449.6   | P79392 | CHROMOGRANIN A.         | 2.26e+01 |
| 43 | 43 | 64.2 | 578.10  | Q04020 | HYPOTHETICAL 66.6 KD P  | 2.26e+01 |
| 44 | 43 | 64.2 | 598.14  | Q66043 | LMPI.                   | 2.26e+01 |
| 45 | 43 | 64.2 | 944.10  | Q48538 | RBOHAP108.              | 2.26e+01 |

```
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NAROO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RL DNA RES. 3:109-136(1996).
DR EMBL: D90910; G1652997; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 414 AA; 44810 MW; 0B3A0E9E CRC32;

Query Match 71.6%; Score 48; DB 2; Length 414;
Best Local Similarity 60.0%; Pred. No. 1.74e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 318 EGPTEAQQQ 327
QY 1 DGPTGEPOQE 10
||| |||:
1 DGPTGEPOQE 10

RESULT 3
ID P79765 PRELIMINARY; PRT; 357 AA.
AC P79765;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE NEUROD-LIKE PROTEIN.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC GALLIFORMES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEGHORN; TISSUE-RETINA;
RA ROZTOCIL T., MATTER-SADZINSKI L., ALLIOD C., BALLIVET M., MATTER J.M.;
RL DEVELOPMENT 124:3263-3272(1997).
DR EMBL: Y09596; E283389; -.
DR PFM: PF00010; HLH.
SQ SEQUENCE 357 AA; 38809 MW; 04A96823 CRC32;

Query Match 70.1%; Score 47; DB 13; Length 357;
Best Local Similarity 66.7%; Pred. No. 2.95e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 9 GPAGEPAQ 17
QY 2 GPTGEPOQE 10
||| |||:
2 GPTGEPOQE 10

RESULT 4
ID 073826 PRELIMINARY; PRT; 357 AA.
AC 073826;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE NEUROD.
GN NEUROD.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC GALLIFORMES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEGHORN;
RA YAN R.-T., WANG S.-Z.;
RL J. NEUROBIOL. 0:0-0(1998).
DR EMBL: AF060885; G3094020; -.
SQ SEQUENCE 357 AA; 38858 MW; 7D406FF3 CRC32;

Query Match 70.1%; Score 47; DB 13; Length 357;
Best Local Similarity 66.7%; Pred. No. 2.95e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 9 GPAGEPAQ 17
```

```
QY 2 GPTGEPOQE 10
||| |||:
2 GPTGEPOQE 10

RESULT 5
ID 044431 PRELIMINARY; PRT; 676 AA.
AC 044431;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE CIRCADIAN CLOCK PROTEIN (FRAGMENT).
GN TIM.
OS DROSOPHILA HYDEI (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN [1]
RP SEQUENCE FROM N.A.
RA OUSLEY A., ZAFARULLAH K., CHEN Y., EMERSON M., HICKMAN L., SEHGAL A.;
RL GENETICS 0:0-0(1998).
DR EMBL: AF038579; G2734130; -.
FT NON_TER 1
FT NON_TER 676
SQ SEQUENCE 676 AA; 76022 MW; 83E96D47 CRC32;

Query Match 70.1%; Score 47; DB 5; Length 676;
Best Local Similarity 60.0%; Pred. No. 2.95e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 357 DGPGKQHQ 366
QY 1 DGPTGEPOQE 10
||| |||:
1 DGPTGEPOQE 10

RESULT 6
ID Q20778 PRELIMINARY; PRT; 1120 AA.
AC Q20778;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE SIMILAR TO TRIPLE HELICAL REGION OF COLLAGENS.
GN F54D8.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 0:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA BENTLEY D.;
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U12966; G529221; -.
SQ SEQUENCE 1120 AA; 119368 MW; E0352B8E CRC32;

Query Match 70.1%; Score 47; DB 5; Length 1120;
Best Local Similarity 60.0%; Pred. No. 2.95e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

Db 1020 EGPTGPGAD 1029  
:||||| :  
QY 1 DGPTGEPQOE 10

RESULT 7  
ID O4430 PRELIMINARY; PRT; 1194 AA.  
AC O4430;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE CIRCADIAN CLOCK PROTEIN (FRAGMENT).  
GN TIM.  
OS DROSOPHILA VIRILIS (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA OUSLEY A., ZAFARULLAH K., CHEN Y., EMERSON M., HICKMAN L., SERGAL A.;  
RL GENETICS 0:0-0(1998).  
DR EMBL; AF038502; G2746733; -.  
FT NON\_TER 1194 1194  
SQ SEQUENCE 1194 AA; 134946 MW; F6C4345B CRC32;

Query Match 70.18; Score 47; DB 5; Length 1194;  
Best Local Similarity 60.0%; Pred. No. 2.95e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Db 486 DGPOGKPOHQ 495  
:||||| :  
QY 1 DGPTGEPQOE 10

RESULT 8  
ID O17482 PRELIMINARY; PRT; 1343 AA.  
AC O17482;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE TIMELESS (TIM).  
GN TIM.  
OS DROSOPHILA VIRILIS (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98033379.  
RA MYERS M.P., ROTHENFLUH A., CHANG M., YOUNG M.W.;  
RL NUCLEIC ACIDS RES. 25:4710-4714(1997).  
DR EMBL; AF032403; G2641617; -.  
DR EMBL; AF032402; G2641617; JOINED.  
SQ SEQUENCE 1343 AA; 150942 MW; 77573CAC CRC32;

Query Match 70.18; Score 47; DB 5; Length 1343;  
Best Local Similarity 60.0%; Pred. No. 2.95e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Db 486 DGPOGKPOHQ 495  
:||||| :  
QY 1 DGPTGEPQOE 10

RESULT 9  
ID O44380 PRELIMINARY; PRT; 1398 AA.  
AC O44380;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE CIRCADIAN CLOCK PROTEIN.  
GN TIM.  
OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98033379.

RA MYERS M.P., ROTHENFLUH A., CHANG M., YOUNG M.W.;  
RL NUCLEIC ACIDS RES. 25:4710-4714(1997).  
DR EMBL; AF032401; G2655282; -.  
DR EMBL; AF032400; G2655282; JOINED.  
SQ SEQUENCE 1398 AA; 156366 MW; 59A88D49 CRC32;

Query Match 70.1%; Score 47; DB 5; Length 1398;  
Best Local Similarity 60.0%; Pred. No. 2.95e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Db 519 DGPOGKPOHQ 528  
:||||| :  
QY 1 DGPTGEPQOE 10

RESULT 10  
ID O35966 PRELIMINARY; PRT; 231 AA.  
AC O35966;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE SEX DETERMINING PROTEIN.  
GN SRI.  
OS MUS MUSCULUS DOMESTICUS (WESTERN EUROPEAN HOUSE MOUSE).  
OC EUKARYOTAE; MITOCHONDRIAL EUKARYOTES; METAZOA; VERTEBRATA;  
OC MAMMALIA; EUTHERIA; RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS;  
OC MUS MUSCULUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-B6JEI-YBUB, CD-1 AND B6JEI-YKAM;  
RX MEDLINE; 98043417.  
RA ALBRECHT K.H., EICHER E.M.;  
RL GENETICS 147:1267-1277(1997).  
DR EMBL; AF009519; G2271479; -.  
DR EMBL; U70643; G2623351; -.  
DR EMBL; U70644; G2623353; -.  
DR PFAM; PF00505; HMG\_box.  
SQ SEQUENCE 231 AA; 28533 MW; 66F6EDC1 CRC32;

Query Match 68.7%; Score 46; DB 11; Length 231;  
Best Local Similarity 70.0%; Pred. No. 4.96e+00;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Db 137 DIPTGHPQOQ 146  
:||||| :  
QY 1 DGPTGEPQOE 10

RESULT 11  
ID O35968 PRELIMINARY; PRT; 232 AA.  
AC O35968;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE SEX DETERMINING PROTEIN.  
GN SRI.  
OS MUS MUSCULUS DOMESTICUS (WESTERN EUROPEAN HOUSE MOUSE).  
OC EUKARYOTAE; MITOCHONDRIAL EUKARYOTES; METAZOA; VERTEBRATA;  
OC MAMMALIA; EUTHERIA; RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS;  
OC MUS MUSCULUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-B6JEI-YWSB, B6JEI-YAPP AND B6JEI-YAKR;  
RX MEDLINE; 98043417.  
RA ALBRECHT K.H., EICHER E.M.;  
RL GENETICS 147:1267-1277(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-B6-YAKR; SPECIES-WESTERN EUROPEAN HOUSE MOUSE;  
RX MEDLINE; 96207297.  
RA CARLSLE C., WINKING H., WEICHENHAN D., NAGAMINE C.M.;  
RL GENOMICS 33:32-45(1996).  
RN [3]



RP SEQUENCE FROM N.A.  
RC STRAIN=B6-YARR; SPECIES=WESTERN EUROPEAN HOUSE MOUSE;  
RA CARLISLE C., NAGAMINE C.M.;  
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; U70650; G2623365; -;  
DR EMBL; U70641; G2623347; -;  
DR EMBL; U70642; G2623349; -;  
DR EMBL; AF068053; G3176975; -;  
DR PFAM; PF00505; HMG\_box  
SQ SEQUENCE 232 AA; 28790 MW; 6998BFF4 CRC32;

Query Match 68.7%; Score 46; DB 11; Length 232;  
Best Local Similarity 70.0%; Pred. No. 4.96e+00;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 137 DIPTGHPQQ 146  
| ||| |||  
QY 1 DGPTGEPOE 10

RESULT 12  
ID O35860 PRELIMINARY; PRT; 233 AA.  
AC O35860;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE SEX DETERMINING PROTEIN.  
GN SRY.  
OS MUS MUSCULUS DOMESTICUS (WESTERN EUROPEAN HOUSE MOUSE).  
OC EUKARYOTAE; MITOCHONDRIAL EUKARYOTES; METAZOA; CHORDATA; VERTEBRATA;  
OC MAMMALIA; EUTHERIA; RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS;  
OC MUS MUSCULUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WMP/EI;  
RX MEDLINE; 98043417.  
RA ALBRECHT K.H.; EICHER E.M.;  
RL GENETICS 147:1267-1277(1997).  
DR EMBL; U70649; G2623363; -;  
DR PFAM; PF00505; HMG\_box.  
SQ SEQUENCE 233 AA; 28790 MW; 3F2451C1 CRC32;

Query Match 68.7%; Score 46; DB 11; Length 233;  
Best Local Similarity 70.0%; Pred. No. 4.96e+00;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 137 DIPTGHPQQ 146  
| ||| |||  
QY 1 DGPTGEPOE 10

RESULT 13  
ID O35858 PRELIMINARY; PRT; 234 AA.  
AC O35858;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE SEX DETERMINING PROTEIN.  
GN SRY.  
OS MUS MUSCULUS DOMESTICUS (WESTERN EUROPEAN HOUSE MOUSE).  
OC EUKARYOTAE; MITOCHONDRIAL EUKARYOTES; METAZOA; CHORDATA; VERTEBRATA;  
OC MAMMALIA; EUTHERIA; RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS;  
OC MUS MUSCULUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B6J1-YLIP;  
RX MEDLINE; 98043417.  
RA ALBRECHT K.H.; EICHER E.M.;  
RL GENETICS 147:1267-1277(1997).  
DR EMBL; U70645; G2623355; -;  
DR PFAM; PF00505; HMG\_box.  
SQ SEQUENCE 234 AA; 28918 MW; 2446EF1B CRC32;

Query Match 68.7%; Score 46; DB 11; Length 234;  
Best Local Similarity 70.0%; Pred. No. 4.96e+00;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 137 DIPTGHPQQ 146  
| ||| |||  
QY 1 DGPTGEPOE 10

RESULT 14  
ID O35859 PRELIMINARY; PRT; 234 AA.  
AC O35859;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE SEX DETERMINING PROTEIN.  
GN SRY.  
OS MUS MUSCULUS DOMESTICUS (WESTERN EUROPEAN HOUSE MOUSE).  
OC EUKARYOTAE; MITOCHONDRIAL EUKARYOTES; METAZOA; CHORDATA; VERTEBRATA;  
OC MAMMALIA; EUTHERIA; RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS;  
OC MUS MUSCULUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B6J1-YORB;  
RX MEDLINE; 98043417.  
RA ALBRECHT K.H.; EICHER E.M.;  
RL GENETICS 147:1267-1277(1997).  
DR EMBL; U70646; G2623357; -;  
DR PFAM; PF00505; HMG\_box.  
SQ SEQUENCE 234 AA; 28918 MW; 69C12948 CRC32;

Query Match 68.7%; Score 46; DB 11; Length 234;  
Best Local Similarity 70.0%; Pred. No. 4.96e+00;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 137 DIPTGHPQQ 146  
| ||| |||  
QY 1 DGPTGEPOE 10

RESULT 15  
ID O69004 PRELIMINARY; PRT; 238 AA.  
AC O69004;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE FERRISIDEROPHORE RECEPTOR-LIKE PROTEIN.  
GN PIGC.  
OS PSEUDOMONAS AERUGINOSA.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
OC PSEUDOMONADACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PAOI;  
RX MEDLINE; 96210657.  
RA OCHSNER U.A.; VASIL M.L.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 93:4409-4414(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PAOI;  
RA OCHSNER U.A.; JOHNSON Z.; VASIL A.I.; VASIL M.L.;  
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AF060193; G3091148; -;  
SQ SEQUENCE 238 AA; 25049 MW; 582EDF82 CRC32;

Query Match 68.7%; Score 46; DB 2; Length 238;  
Best Local Similarity 55.6%; Pred. No. 4.96e+00;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 182 GSTGDPQRD 190  
|:|:|:|:|  
QY 2 GPTGEPOE 10

Search completed: Thu May 13 15:37:55 1999  
Job time : 34 secs.

\*\*\*\*\*  
M P E R F A  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 13 15:39:59 1999; MasPar time 2.44 Seconds  
Tabular output not generated. 39.713 Million cell updates/sec

Title: >US-09-040-485-8  
Description: (3-8) from US09040485.pep (2 of 2)  
Perfect Score: 41  
Sequence: 1 PTGEPQ 6

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq32  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 12.609; Variance 30.187; scale 0.418

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES  |       |             |                |                       |
|------------|-------|-------------|----------------|-----------------------|
| Result No. | Score | Query Match | Description    | Pred. No.             |
| 1          | 38    | 92.7        | 539 14 R80506  | S. lividans protease  |
| 2          | 37    | 90.2        | 3038 14 R74171 | Aspergillus terreus t |
| 3          | 35    | 85.4        | 512 3 R14908   | ORF from BIK1-FUS1 re |
| 4          | 35    | 85.4        | 512 1 P94880   | FUS 1 promoter incorp |
| 5          | 35    | 85.4        | 512 3 R14909   | FUS1 protein.         |
| 6          | 35    | 85.4        | 1931 5 R27649  | Human calcium channel |
| 7          | 35    | 85.4        | 2237 6 R33550  | Sequence of the alpha |
| 8          | 35    | 85.4        | 2237 14 R71006 | Human neuronal calciu |
| 9          | 35    | 85.4        | 2339 6 R33549  | Sequence of the alpha |
| 10         | 35    | 85.4        | 2339 14 R71005 | Human neuronal calciu |
| 11         | 34    | 82.9        | 59 25 W32784   | Rat thymus 60K-CSF pr |
| 12         | 34    | 82.9        | 109 20 W02134  | High mobility group H |
| 13         | 34    | 82.9        | 177 16 R81430  | Hepatitis GB virus (H |
| 14         | 34    | 82.9        | 343 20 W02132  | High mobility group p |
| 15         | 34    | 82.9        | 649 29 W38218  | Maize starch synthase |
| 16         | 34    | 82.9        | 659 20 W06708  | Mouse haematopoietic- |
| 17         | 34    | 82.9        | 680 2 R07433   | First open reading fr |
| 18         | 34    | 82.9        | 757 20 W03179  | Bovine poly-immunoglo |

|    |    |      |                |                       |
|----|----|------|----------------|-----------------------|
| 19 | 34 | 82.9 | 806 5 R27481   | RP-III residual prote |
| 20 | 34 | 82.9 | 1403 10 R54629 | Mannuronan C-5-epimer |
| 21 | 34 | 82.9 | 1838 10 R53257 | Human collagen (type  |
| 22 | 34 | 82.9 | 2322 22 W15566 | Melanoma-associated c |
| 23 | 34 | 82.9 | 3567 8 R44431  | eryA region polypepti |
| 24 | 34 | 80.5 | 46 24 W16397   | HSV-1 glycoprotein B  |
| 25 | 33 | 80.5 | 167 4 R22273   | Human collagen XI ins |
| 26 | 33 | 80.5 | 400 2 P70654   | Sequence encoded by g |
| 27 | 33 | 80.5 | 442 20 W06709  | Human haematopoietic- |
| 28 | 33 | 80.5 | 454 4 R25155   | Fibronectin-spacer-co |
| 29 | 33 | 80.5 | 478 9 R47586   | Human bone formation- |
| 30 | 33 | 80.5 | 478 20 W06539  | Human bone morphogen  |
| 31 | 33 | 80.5 | 579 20 W08143  | RHDV capsid protein.  |
| 32 | 33 | 80.5 | 659 17 R94534  | BTK tyrosine kinase.  |
| 33 | 33 | 80.5 | 903 2 P70426   | Recombinant herpes si |
| 34 | 33 | 80.5 | 903 3 P60244   | Herpes simplex virus  |
| 35 | 33 | 80.5 | 903 3 P71135   | Herpes simplex virus- |
| 36 | 33 | 80.5 | 904 8 R41779   | Glycoprotein B (gB2). |
| 37 | 33 | 80.5 | 904 20 W00375  | HSV-1 glycoprotein B. |
| 38 | 33 | 80.5 | 904 26 W34553  | Herpes simplex virus  |
| 39 | 33 | 80.5 | 905 3 P80915   | Sequence of Herpes si |
| 40 | 33 | 80.5 | 1058 9 R42453  | Enzyme involved in ei |
| 41 | 33 | 80.5 | 1366 13 R17102 | Collagen alpha 2 (I)  |
| 42 | 33 | 80.5 | 2756 20 R99462 | Biosynthetic enzyme o |
| 43 | 33 | 80.5 | 3080 1 P93285  | Sequence of clone HIV |
| 44 | 33 | 80.5 | 3210 1 P81771  | Deduced sequence enco |
| 45 | 32 | 78.0 | 497 17 R81462  | Human derived cytochr |

ALIGNMENTS

RESULT 1  
ID R80506 standard; Protein; 539 AA.  
AC R80506;  
DT 04-DEC-1995 (first entry)  
DE S. lividans protease P5-6.  
KW Protease: metalloendoproteinase; tripeptidyl aminopeptidase;  
OS Protease-deficiency; protein secretion.  
OS Streptomyces lividans.  
FH Key Location/Qualifiers  
FT peptide 1..47  
FT /label= Sig\_peptide  
PN W09517512-A.  
PD 29-JUN-1995.  
PF 22-DEC-1994; U14772.  
PR 23-DEC-1993; US-173508.  
PA (CANG-) CANGENE CORP.  
PI Bartfeld D, Butler MJ, Hadary D, Jenish DL, Krieger TJ;  
PI Malek LT, Soostmeyer G, Walczyk E;  
DR WPI; 95-240673/31.  
DR N-PSDB; Q99366.  
PT Endogenous Streptomyces protease(s), opt. having impaired activity -  
PT useful in prodn. of exogenous proteins with reduced proteolytic  
PT degradation  
PS Claim 4; Fig.14; 142pp; English.  
CC Protease Tap-negative cells were transformed with a S. lividans 66  
CC genomic library and screened with APA-beta-naphthylamide to  
CC isolate colonies contg. genes (Q99365-68) for novel proteases  
CC P5-4, P5-6, P5-10 and P8-2 (R80505-08). Impaired expression of  
CC such proteases by Streptomyces hosts improves the quality, quantity  
CC and stability of exogenous gene products.  
SQ Sequence 539 AA;

Query Match 92.7%; Score 38; DB 14; Length 539;  
Best Local Similarity 83.3%; Pred No. 8.16e+01;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 457 ptgeph 462  
QY 3 PTGEPQ 8  
RESULT 2

```

ID R74171 standard; Protein; 3038 AA.
AC R74171;
DE 18-JAN-1996 (first entry)
DE Aspergillus terreus triol polyketide synthase.
KW triol polyketide synthase; TPKS; HMG-CoA reductase inhibitor;
KW hypercholesterolaemia; LDL-cholesterol.
OS Aspergillus terreus.
FH Key Location/Qualifiers
FT region 181
FT /label= keto-acyl synthase motif
FT region 654..658
FT /label= acetyl/malonyl transferase motif
FT region 985..994
FT /label= dehydratase motif
FT region 1446..1450
FT /label= methyl transferase motif
FT region 1932..1937
FT /label= enoyl reductase motif
FT region 2164..2169
FT /label= keto reductase motif
FT region 2498
FT /label= acyl carrier protein motif
FT misc_difference 282..288
FT /label= misc feature
FT misc_difference 1450..1460
FT /label= misc feature
FT misc_difference 1603..1612
FT /label= misc feature
FT misc_difference 2521..2535
FT /label= misc feature
FT WO9512661-A.
PN 11-MAY-1995.
PD 28-OCT-1994; U12423.
PF 02-NOV-1993; US-148132.
PR (MERI) MERCK & CO INC.
PA Conder MJ, Davis CR, Hendrickson LE, Mcada PC, Rambosek J;
PI Reeves CD, Vinci VA;
DR WPI: 95-193816/25.
DR N-PSDB; Q92323.
DR Novel DNA encoding triol poly-ketide synthase - used to isolate and
PT identify homologues of triol poly-ketide synthase, and in the treatment
PT of hyper-cholesterolaemia
PS Claim 12; Figure 2; 107pp; English.
CC The full-length TPKS-encoding DNA in plasmid pLOA was
CC designated pPKS100. Splicing of the introns from the DNA
CC sequence and translation of the 914 nt ORF results in a
CC protein of 3038 AAs (R74171) with a mol. wt. of 269,090
CC daltons. Inspection of the TPKS AA sequence for active
CC site residues and motifs known to be associated with
CC polyketide synthases and fatty acid synthase (FAS) activities
CC resulted in the identification of candidates for expected
CC sites (see FT). Except for the presence of a methyl transferase,
CC not present in FAS, the succession of activities on the
CC TPKS protein is the same as that observed for the rat FAS
CC protein.
CC Sequence 3038 AA;
SQ
Query Match 90.2%; Score 37; DB 14; Length 3038;
Best Local Similarity 66.7%; Pred. No. 1,11e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 325 pagdpq 330
|:|:|
QY 3 PTGEPQ 8

RESULT 3
ID R14908 standard; Protein; 512 AA.
AC R14908;
DE 03-FEB-1992 (first entry)
DE ORF from BIK1-FUS1 region.
KW Pheromone inducible yeast promoter; bilateral karyogamy defect;
KW FUS2; BIK1.

OS Saccharomyces cerevisiae.
PN US5063154-A.
PD 05-NOV-1991.
PF 24-JUN-1988; 212270.
PR 24-JUN-1987; US-066078.
PR 24-JUN-1988; US-212270.
PA (WHIT-) WHITEHEAD INST BIOM.
PI Fink GR, Trueheart J, Elion EA;
DR WPI: 91-346534/47.
DR N-PSDB; Q14754.
DR DNA fragment contg. pheromone-inducible yeast promoter - useful
PT for transforming yeast cells to produce foreign proteins, which
PT may be toxic to yeast cells.
PS Disclosure; Fig 4; 23pp; English.
CC The ORF occurs between the BIK1 reading frame (on the complementary
CC strand) and the FUS1 reading frame on chromosome III of S. cerevis-
CC iae. Transcription of the FUS1 gene is greatly enhanced by the
CC presence of the appropriate mating pheromone. The promoter region
CC can be used for the pheromone inducible expression of proteins
CC of interest.
CC See also R14907-10.
CC Sequence 512 AA;
SQ
Query Match 85.4%; Score 35; DB 3; Length 512;
Best Local Similarity 83.3%; Pred. No. 2.06e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 282 ptgapq 287
|:|:|
QY 3 PTGEPQ 8

RESULT 4
ID P94880 standard; Protein; 512 AA.
AC P94880;
DE 02-JUL-1990 (first entry)
DE FUS 1 promoter incorporated within the HIS4 gene.
DE Yeast promoter; yeast pheromones; FUS-1; BIK-1; HIS-4.
OS Saccharomyces cerevisiae.
PN WO8810308-A.
PD 29-DEC-1988.
PF 23-JUN-1988; 02129.
PF 24-JUN-1987; US-066078.
PA (WHIT-) Whitehead Inst.
PI Fink GR, Trueheart J, Elion EA;
DR WPI: 89-023850/03.
DR N-PSDB; N93100.
DR New DNA fragment contg. protein encoding gene and yeast promoter -
PT controlled by mating pheromone allowing efficient and
PT regulatable expression.
PS Claim 5; Fig 4; 51pp; English.
CC FUS 1 promoter found within HIS4 gene may be used to promote at least one
CC polypeptide within a high-copy vector induced by a-factor for alpha cells
CC and alpha-factor for a-cells.
CC Sequence 512 AA;
SQ
Query Match 85.4%; Score 35; DB 1; Length 512;
Best Local Similarity 83.3%; Pred. No. 2.06e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 282 ptgapq 287
|:|:|
QY 3 PTGEPQ 8

RESULT 5
ID R14909 standard; Protein; 512 AA.
AC R14909;
DE 03-FEB-1992 (first entry)
DE FUS1 protein.
KW Pheromone inducible yeast promoter; bilateral karyogamy defect;
KW FUS2; BIK1.
OS Saccharomyces cerevisiae.

```

PN US063154-A.  
 PD 05-NOV-1991.  
 PF 24-JUN-1988; 212270.  
 PR 24-JUN-1987; US-066078.  
 PR 24-JUN-1988; US-212270.  
 PA (WHT-) WHITEHEAD INST BIOM.  
 PI Fink GR, Trueheart J, Eilion EA;  
 DR WPI; 91-346534/47.  
 DR N-PSDB; Q14754.  
 PT DNA fragment contg. pheromone-inducible yeast promoter - useful  
 PT for transforming yeast cells to produce foreign proteins, which  
 PT may be toxic to yeast cells.  
 PS Disclosure; Fig 4; 23pp; English.  
 CC The FUS1 gene is involved in the initial zygote formation (cell  
 CC surface recognition leading to cytoplasmic fusion) during conju-  
 CC gation. Transcription of the FUS1 gene is greatly enhanced by the  
 CC presence of the appropriate mating pheromone. The promoter region  
 CC can be used for the pheromone inducible expression of proteins  
 CC of interest.  
 CC See also RI4907-10.  
 CC Sequence 512 AA;  
 SQ

Query Match 85.4%; Score 35; DB 3; Length 512;  
 Best Local Similarity 83.3%; Pred. No. 2.06e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 282 ptgapq 287  
 |||||  
 QY 3 PTGEPQ 8

## RESULT 6

ID R27649 standard; Protein; 1931 AA.  
 AC R27649.  
 DT 03-MAR-1993 (first entry)  
 DE Human calcium channel 27980/11.  
 KW Plasmid pRI4-5.3.3.1; Ca-flux assay.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_difference 1165  
 FT /note= "encoded by GNG codon, N is unknown"

PN EP-507170-A.  
 PD 07-OCT-1992.  
 PF 23-MAR-1992; 104970.  
 PR 04-APR-1991; DE-110785.  
 PA (FARB ) BAYER AG.  
 PI Franz J, Rae P, Unterbeck A, Weingaertner B;  
 DR WPI; 92-333446/41.  
 DR N-PSDB; Q29269.  
 PT Cloned human neuronal calcium channel sub-types - useful in  
 PT calcium flux assays to screen for neurone-specific calcium  
 PT channel ligands  
 PS Claim 2; Page 63-77; 101pp; German.  
 CC Human neuroblastoma cell line, hippocampus, frontal and temporal  
 CC cortex and visual cortex cDNA banks were screened with a probe  
 CC containing carp skeletal muscle Ca-channel cDNA. The cDNA clone  
 CC pRI4-5.3.3.1 overlaps with clone p1247-14.1.1.1 (see Q29263).  
 CC There were a number of differences between the two sequences  
 CC including the deletion of an Adenosine residue at position 1013 of  
 CC p1247-14.1.1.1 which leads to a stop codon at position 1028-1030;  
 CC the deletion is thought to be a cloning artefact. The human  
 CC neuronal calcium channel protein can be used for screening for Ca  
 CC channel ligands (agonists or antagonists). See also Q29259-Q29275.  
 SQ Sequence 1931 AA;

Query Match 85.4%; Score 35; DB 5; Length 1931;  
 Best Local Similarity 83.3%; Pred. No. 2.06e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1577 pdgepq 1582  
 |||||  
 QY 3 PTGEPQ 8

## RESULT 7

ID R33550 standard; Protein; 2237 AA.  
 AC R33550;  
 DT 30-JUN-1993 (first entry)  
 DE Sequence of the alpha 1B-2 human calcium channel subunit.  
 KW Human calcium channel subunit; diagnosis; agonist; antagonist;  
 KW Lambert Eaton syndrome.  
 OS Homo sapiens.  
 PN W09304083-A.  
 PD 04-MAR-1993.  
 PF 14-AUG-1992; U06903.  
 PR 15-AUG-1991; US-745206.  
 PR 10-APR-1992; US-868354.  
 PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, Mccue AF,  
 PI Williams ME;  
 DR WPI; 93-093936/11.  
 DR N-PSDB; Q37818.  
 PT DNA encoding specific human calcium channel sub-units - used for  
 PT identifying calcium channel agonists and antagonists and  
 PT diagnosing Lambert Eaton syndrome

PS Disclosure; Page 120-128; 150pp; English.  
 CC DNA encoding the alpha 1B subunit was isolated by screening a  
 CC human basal ganglia cDNA library with fragments of the rabbit  
 CC skeletal muscle calcium channel alpha 1 subunit-encoding cDNA.  
 CC A portion of one of the positive clones was used to screen an IMR32  
 CC cell cDNA library. Clones that hybridized to the basal ganglia  
 CC DNA prove were used to further screen an IMR32 cell cDNA library  
 CC to identify overlapping clones that in turn were used to screen a  
 CC human hippocampus cDNA library. In this way, a sufficient series of  
 CC clones to span nearly the entire length of the nucleotide sequence  
 CC encoding the human alpha 1B subunit was obtained. PCR amplification  
 CC of specific regions of the IMR32 cell alpha 1B mRNA yielded  
 CC additional segments of the alpha 1B coding sequence. A full-length  
 CC alpha 1B DNA clone was constructed by ligating portions of the  
 CC partial cDNA clones (see Q37817, Q37818). Alpha 1B-1 and alpha  
 CC 1B-2 are derived by alternative splicing of the alpha 1B subunit  
 CC transcript.  
 SQ Sequence 2237 AA;

Query Match 85.4%; Score 35; DB 6; Length 2237;  
 Best Local Similarity 83.3%; Pred. No. 2.06e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1985 pdgepq 1990  
 |||||  
 QY 3 PTGEPQ 8

## RESULT 8

ID R71006 standard; Protein; 2237 AA.  
 AC R71006;  
 DT 01-DEC-1995 (first entry)  
 DE Human neuronal calcium channel subunit alpha 1B-2.  
 KW Calcium channel subunit; antagonist; agonist; diagnosis;  
 KW Lambert Eaton Syndrome.  
 OS Homo sapiens.  
 PN W09504822-A.  
 PD 16-FEB-1995.  
 PF 11-AUG-1994; U09230.  
 PR 11-AUG-1993; US-105536.  
 PR 05-NOV-1993; US-149097.  
 PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
 PI Ellis SB, Gillespie A, Harpold MM, Mccue AF, Williams ME;  
 DR WPI; 95-090900/12.  
 DR N-PSDB; Q84658.  
 PT DNA encoding human calcium channel sub-unit(s) - used for  
 PT developing prods. for studying calcium channels, e.g. for  
 PT obtaining agonists and antagonists

PS Disclosure; Page 149-160; 285pp; English.  
 CC DNA encoding the alpha 1B subunit was isolated by screening a  
 CC human basal ganglia cDNA library with fragments of the rabbit

CC skeletal muscle calcium channel alpha 1 subunit-encoding cDNA.  
 CC A portion of one of the positive clones was used to screen an  
 CC IMR32 cell cDNA library. Clones that hybridized to the basal  
 CC ganglia probe were used to further screen an IMR32 cell cDNA  
 CC library to identify overlapping clones that in turn were used  
 CC to screen a human hippocampus cDNA library. A series of clones  
 CC to span nearly the entire length of the nt. sequence encoding  
 CC the human alpha 1B subunit was obtd. Nucleic acid amplification  
 CC of specific regions of the IMR32 cell alpha 1B mRNA yielded  
 CC additional segments of the alpha 1B coding sequence. A full-  
 CC length alpha 1B DNA clone was constructed by ligating portions  
 CC of the partial cDNA clones. Nucleic acid amplification analysis  
 CC of IMR32 cell RNA and genomic DNA using oligo primers corresp. to  
 CC sequences located 5' and 3' of the stop codon of the DNA encoding  
 CC the alpha 1B subunit revealed an alternatively spliced alpha  
 CC 1B-encoding mRNA in IMR32 cells. This second mRNA product is the  
 CC result of differential splicing of the alpha 1B subunit transcript  
 CC to include another exon that is not present in the mRNA corresp.  
 CC to the other 3' alpha 1B cDNA sequence that was initially isolated.  
 CC The alpha 1B subunit encoded by a DNA sequence contg. an additional  
 CC exon is referred to as alpha 1B-1 and given in Q84657/R71005,  
 CC whereas the other form is referred to as alpha 1B-2 and is given in  
 CC Q84658/R71006. Following the sequence of the additional exon in  
 CC alpha 1B-1 the alpha 1B-1 and alpha 1B-2 sequences are identical.  
 CC Sequence 2237 AA;

Query Match 85.4%; Score 35; DB 14; Length 2237;  
 Best Local Similarity 83.3%; Pred. No. 2.06e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1985 pdgepq 1990  
 Qy 3 PTGEPQ 8

RESULT 9  
 ID R33549 standard; Protein; 2339 AA.

DE 30-JUN-1993 (first entry)  
 DT Sequence of the alpha 1B-1 human calcium channel subunit.  
 KW Human calcium channel subunit; diagnosis; agonist; antagonist;  
 KW Lambert Eaton syndrome.  
 OS Homo sapiens.  
 PN W09304083-A.

PD 04-MAR-1993.  
 PF 14-AUG-1992; U06903.  
 PR 15-AUG-1991; US-745206.  
 PR 10-APR-1992; US-868354.  
 PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, Mccue AF,  
 PI Williams ME;  
 DR WPI; 93-093936/11.  
 DR N-PSDB; Q37817.

PT DNA encoding specific human calcium channel sub-units - used for  
 PT identifying calcium channel agonists and antagonists and  
 PT diagnosing Lambert Eaton syndrome  
 PS Disclosure; Page 111-120; 150pp; English.  
 CC DNA encoding the alpha 1B subunit was isolated by screening a  
 CC human basal ganglia cDNA library with fragments of the rabbit  
 CC skeletal muscle calcium channel alpha 1 subunit-encoding cDNA.  
 CC A portion of one of the positive clones was used to screen an IMR32  
 CC cell cDNA library. Clones that hybridized to the basal ganglia  
 CC DNA probe were used to further screen an IMR32 cell cDNA library  
 CC to identify overlapping clones that in turn were used to screen a  
 CC human hippocampus cDNA library. In this way, a sufficient series of  
 CC clones to span nearly the entire length of the nucleotide sequence  
 CC encoding the human alpha 1B subunit was obtained. PCR amplification  
 CC of specific regions of the IMR32 cell alpha 1B mRNA yielded  
 CC additional segments of the alpha 1B coding sequence. A full-length  
 CC alpha 1B DNA clone was constructed by ligating portions of the  
 CC partial cDNA clones (see Q37817, Q37818). Alpha 1B-1 and alpha  
 CC 1B-2 are derived by alternative splicing of the alpha 1B subunit  
 CC transcript.

SQ Sequence 2339 AA;

Query Match 85.4%; Score 35; DB 6; Length 2339;  
 Best Local Similarity 83.3%; Pred. No. 2.06e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1985 pdgepq 1990  
 Qy 3 PTGEPQ 8

RESULT 10  
 ID R71005 standard; Protein; 2339 AA.

AC R71005;  
 DT 01-DEC-1995 (first entry)  
 DE Human neuronal calcium channel subunit alpha 1B-1.  
 KW Calcium channel subunit; antagonist; agonist; diagnosis;  
 KW Lambert Eaton Syndrome.  
 OS Homo sapiens.

PN W09504822-A.  
 PD 16-FEB-1995.  
 PF 11-AUG-1994; U09230.  
 PR 11-AUG-1993; US-105536.  
 PR 05-NOV-1993; US-149097.  
 PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
 PI Ellis SB, Gillespie A, Harpold MM, Mccue AF, Williams ME;  
 DR WPI; 95-090900/12.  
 DR N-PSDB; Q84657.

PT DNA encoding human calcium channel sub-unit(s) - used for  
 PT developing prods. for studying calcium channels, e.g. for  
 PT obtaining agonists and antagonists  
 PS Disclosure; Page 138-149; 285pp; English.  
 CC DNA encoding the alpha 1B subunit was isolated by screening a  
 CC human basal ganglia cDNA library with fragments of the rabbit  
 CC skeletal muscle calcium channel alpha 1 subunit-encoding cDNA.  
 CC A portion of one of the positive clones was used to screen an  
 CC IMR32 cell cDNA library. Clones that hybridized to the basal  
 CC ganglia probe were used to further screen an IMR32 cell cDNA  
 CC library to identify overlapping clones that in turn were used  
 CC to screen a human hippocampus cDNA library. A series of clones  
 CC to span nearly the entire length of the nt. sequence encoding  
 CC the human alpha 1B subunit was obtd. Nucleic acid amplification  
 CC of specific regions of the IMR32 cell alpha 1B mRNA yielded  
 CC additional segments of the alpha 1B coding sequence. A full-  
 CC length alpha 1B DNA clone was constructed by ligating portions  
 CC of the partial cDNA clones. Nucleic acid amplification analysis  
 CC of IMR32 cell RNA and genomic DNA using oligo primers corresp. to  
 CC sequences located 5' and 3' of the stop codon of the DNA encoding  
 CC the alpha 1B subunit revealed an alternatively spliced alpha  
 CC 1B-encoding mRNA in IMR32 cells. This second mRNA product is the  
 CC result of differential splicing of the alpha 1B subunit transcript  
 CC to include another exon that is not present in the mRNA corresp.  
 CC to the other 3' alpha 1B cDNA sequence that was initially isolated.  
 CC The alpha 1B subunit encoded by a DNA sequence contg. an additional  
 CC exon is referred to as alpha 1B-1 and given in Q84657/R71005,  
 CC whereas the other form is referred to as alpha 1B-2 and is given in  
 CC Q84658/R71006. Following the sequence of the additional exon in  
 CC alpha 1B-1 the alpha 1B-1 and alpha 1B-2 sequences are identical.  
 CC Sequence 2339 AA;

Query Match 85.4%; Score 35; DB 14; Length 2339;  
 Best Local Similarity 83.3%; Pred. No. 2.06e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1985 pdgepq 1990  
 Qy 3 PTGEPQ 8

RESULT 11  
 ID W32784 standard; Protein; 59 AA.

AC W32784;  
 DT 03-FEB-1998 (first entry)



FT misc\_difference 302 /note= "lipoma preferred partner fragment"  
 FT /note= "corresponding codon TAG"  
 FT misc\_difference 320 /note= "corresponding codon TAA"  
 FT misc\_difference 325 /note= "corresponding codon TAA"  
 FT EP-727487-A1.  
 FN 21-AUG-1996.  
 PD 14-JUL-1995; 201951.  
 PR 17-FEB-1995; EP-200390.  
 PA (LEUV-) LEUVEN RES & DEV.  
 PI Bullerdiek J, Moles R, Schoenmakers HPPM, Van De Ven WJM;  
 DR WPI; 96-372875/38.  
 DR N-PSDB; T36363.  
 PT Multi-tumour aberrant growth gene - and probes derived from it,  
 PT useful in in situ diagnostic method for detecting tumour cells, e.g.  
 PT mesenchymal hamartomas and rhabdomyosarcomas  
 FS Example 3; Pages 45-47; 78pp; English.  
 CC The present sequence is the predicted protein product of a 3'-RACE  
 CC product, comprising the junction region of a high mobility group  
 CC protein (HMG):lipoma preferred partner (LPP) fusion gene. The  
 CC protein product was predicted to contain 3 DNA binding domains  
 CC (DBD) of the HMG1-C (map position 12q15), and the 3 cysteine rich  
 CC zinc binding (LIM) domains of the LPP (map position 3q27-28). In  
 CC PCR analysis using appropriate nested amplicon sets, similar  
 CC HMG:LPP hybrid transcripts were detected in various primary lipomas  
 CC and lipoma cell lines carrying a chromosome 3 to 12 translocation  
 CC (t(3;12)), and in a cytogenetically normal lipoma. This revealed  
 CC that the cytogenetically detectable, and hidden, t(3;12) in lipomas  
 CC result in the in-phase fusion of the DBD of HMG1-C to the LIM  
 CC domains of LPP. Consequently, these LIM domains are most likely  
 CC presented in the nuclear environment of these lipoma cells, where  
 CC they might affect gene expression, possibly leading to aberrant  
 CC growth control. This in an example of a benign mesenchymal tumour  
 CC with chromosome 12q13-q15 aberrations, where a chromosome  
 CC translocation partner contributes recurrently and consistently to  
 CC the formation of a well-defined tumour associated HMG1-C fusion  
 CC protein.  
 SQ Sequence 343 AA;

Query Match 82.9%; Score 34; DB 20; Length 343;  
 Best Local Similarity 100.0%; Pred. No. 2.80e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 16 ptgep 20  
 |||||  
 Qy 3 PTGEP 7

RESULT 15  
 ID W38218 standard; Protein; 649 AA.  
 AC W38218;  
 DT 22-MAY-1998 (first entry)  
 DE Maize starch synthase type I.  
 KW Maize; starch synthase type I; starch.  
 OS Zea mays.  
 FN DE19619918-AL.  
 PD 20-NOV-1997.  
 PF 17-MAY-1996; 019918.  
 PR 17-MAY-1996; DE-019918.  
 PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.  
 PI Froberg C, Kossmann J;  
 DR WPI; 98-000821/01.  
 DR N-PSDB; T95785.  
 PT DNA encoding maize starch synthase type I protein - for producing  
 PT transgenic plants  
 PS Claim 1; Pages 16-20; 23pp; German.  
 CC The present sequence is maize starch synthase type I, useful  
 CC in the production of starch. Starch can be used in various  
 CC conventional starch applications, e.g. starch hydrolysate products,  
 CC foods, papermaking, adhesives, textiles, building materials, soil  
 CC stabilisation, agrochemicals, pharmaceuticals, cosmetics, coal

CC briquettes, ore and coal slurries, foundry casting, rubber, leather  
 CC and synthetic polymers. The enzyme produces a starch stated to have  
 CC different physicochemical properties, especially viscosity and  
 CC gelling properties, from wild type starch.  
 SQ Sequence 649 AA;

Query Match 82.9%; Score 34; DB 29; Length 649;  
 Best Local Similarity 100.0%; Pred. No. 2.80e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 83 ptgep 87  
 |||||  
 Qy 3 PTGEP 7

Search completed: Thu May 13 15:40:16 1999  
 Job time : 17 secs.



\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 13 15:41:05 1999; MasPar time 1.94 Seconds

Tabular output not generated. 83.113 Million cell updates/sec

Title: >US-09-040-485-8

Description: (3-8) from US09040485.pep (2 of 2)

Perfect Score: 41

Sequence: 1 PTGEPQ 6

Scoring table:

PAM 150

Gap 15

Searched: 74019 seqs, 26840295 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot35

1:swissprot

Statistics: Mean 18.021; Variance 16.071; scale 1.121

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description             | Pred. No. |
|------------|-------|---------------|--------|-------|-------------------------|-----------|
| 1          | 41    | 100.0         | 757    | 1     | ASPH_HUMAN              | 5.09e-01  |
| 2          | 38    | 92.7          | 232    | 1     | ASPARTYL/ASPARAGINYL B  | 3.60e+00  |
| 3          | 38    | 92.7          | 485    | 1     | HYPOTHETICAL GENE 8 ME  | 3.60e+00  |
| 4          | 38    | 92.7          | 518    | 1     | GLUTAMYL-TRNA SYNTHETA  | 3.60e+00  |
| 5          | 38    | 92.7          | 520    | 1     | PUTATIVE CYTOCHROME P4  | 3.60e+00  |
| 6          | 38    | 92.7          | 520    | 1     | PUTATIVE CYTOCHROME P4  | 3.60e+00  |
| 7          | 37    | 90.2          | 116    | 1     | PUTATIVE TYROSINE PHOSP | 6.75e+00  |
| 8          | 37    | 90.2          | 301    | 1     | VP6 PROTEIN (MINOR INN  | 6.75e+00  |
| 9          | 37    | 90.2          | 325    | 1     | VP6 PROTEIN (MINOR INN  | 6.75e+00  |
| 10         | 37    | 90.2          | 325    | 1     | VP6 PROTEIN (MINOR INN  | 6.75e+00  |
| 11         | 37    | 90.2          | 325    | 1     | VP62_BTIV10             | 6.75e+00  |
| 12         | 37    | 90.2          | 325    | 1     | VP62_BTIV13             | 6.75e+00  |
| 13         | 37    | 90.2          | 328    | 1     | VP6 PROTEIN (MINOR INN  | 6.75e+00  |
| 14         | 37    | 90.2          | 329    | 1     | VP6 PROTEIN (MINOR INN  | 6.75e+00  |
| 15         | 37    | 90.2          | 370    | 1     | VP61_BTIV10             | 6.75e+00  |
| 16         | 37    | 90.2          | 468    | 1     | GLUTAMYL-TRNA SYNTHETA  | 6.75e+00  |
| 17         | 36    | 87.8          | 142    | 1     | GENERAL SECRETION PATH  | 1.25e+01  |
| 18         | 36    | 87.8          | 612    | 1     | GSPG_PSEAE              | 1.25e+01  |
| 19         | 36    | 87.8          | 1079   | 1     | LEPA_BACSU              | 1.25e+01  |
| 20         | 35    | 85.4          | 218    | 1     | GTP-BINDING PROTEIN LE  | 1.25e+01  |
| 21         | 35    | 85.4          | 231    | 1     | YD23_SCHPO              | 2.29e+01  |
| 22         | 35    | 85.4          | 251    | 1     | HYPOTHETICAL 119.9 KD   | 2.29e+01  |
| 23         | 35    | 85.4          | 311    | 1     | POU2_XENLA              | 2.29e+01  |
|            |       |               |        |       | TRANSCRIPTION FACTOR P  | 2.29e+01  |
|            |       |               |        |       | RIBONUCLEASE III (EC 3  | 2.29e+01  |
|            |       |               |        |       | PRP2_HUMAN              | 2.29e+01  |
|            |       |               |        |       | SALIVARY PROLINE-RICH   | 2.29e+01  |
|            |       |               |        |       | SEX-DETERMINING REGION  | 2.29e+01  |

|    |    |      |      |   |            |                        |          |
|----|----|------|------|---|------------|------------------------|----------|
| 24 | 35 | 85.4 | 331  | 1 | PRP1_HUMAN | SALIVARY PROLINE-RICH  | 2.29e-01 |
| 25 | 35 | 85.4 | 333  | 1 | CAP1_DICDI | CAMP BINDING PROTEIN C | 2.29e-01 |
| 26 | 35 | 85.4 | 337  | 1 | GALE_SALTY | UDP-GLUCOSE 4-EPIMERAS | 2.29e-01 |
| 27 | 35 | 85.4 | 355  | 1 | SRY_MUSSP  | SEX-DETERMINING REGION | 2.29e-01 |
| 28 | 35 | 85.4 | 512  | 1 | FUS1_YEAST | NUCLEAR FUSION PROTEIN | 2.29e-01 |
| 29 | 35 | 85.4 | 535  | 1 | YZ25_MYCTU | HYPOTHETICAL 56.2 KD P | 2.29e-01 |
| 30 | 35 | 85.4 | 545  | 1 | HMWC_DESVH | HIGH-MOLECULAR-WEIGHT  | 2.29e-01 |
| 31 | 35 | 85.4 | 562  | 1 | O16G_BACTR | OLIGO-1,6-GLUCOSIDASE  | 2.29e-01 |
| 32 | 35 | 85.4 | 618  | 1 | DNAK_DEIPR | DNAK PROTEIN (HEAT SHO | 2.29e-01 |
| 33 | 35 | 85.4 | 622  | 1 | MAK_RAT    | SERINE/THREONINE-PROTE | 2.29e-01 |
| 34 | 35 | 85.4 | 622  | 1 | MAK_MOUSE  | SERINE/THREONINE-PROTE | 2.29e-01 |
| 35 | 35 | 85.4 | 2336 | 1 | CIC5_RAT   | N-TYPE CALCIUM CHANNEL | 2.29e-01 |
| 36 | 35 | 85.4 | 2339 | 1 | CIC5_HUMAN | N-TYPE CALCIUM CHANNEL | 2.29e-01 |
| 37 | 35 | 85.4 | 3712 | 1 | LMA_DROME  | LAMININ ALPHA CHAIN PR | 2.29e-01 |
| 38 | 34 | 82.9 | 296  | 1 | PRP3_MOUSE | PROLINE-RICH PROTEIN M | 4.12e-01 |
| 39 | 34 | 82.9 | 369  | 1 | PSPD_BOVIN | PULMONARY SURFACTANT-A | 4.12e-01 |
| 40 | 34 | 82.9 | 381  | 1 | YE21_CAEEL | HYPOTHETICAL 42.2 KD P | 4.12e-01 |
| 41 | 34 | 82.9 | 475  | 1 | TRB1_ECOLI | TRAB PROTEIN.          | 4.12e-01 |
| 42 | 34 | 82.9 | 491  | 1 | PRCC_HUMAN | PROLINE-RICH PROTEIN P | 4.12e-01 |
| 43 | 34 | 82.9 | 928  | 1 | YL83_CAEEL | HYPOTHETICAL 105.8 KD  | 4.12e-01 |
| 44 | 34 | 82.9 | 1220 | 1 | ATCP_PIG   | CALCIUM-TRANSPORTING A | 4.12e-01 |
| 45 | 34 | 82.9 | 1319 | 1 | MNL_HUMAN  | PROBABLE TUMOR SUPPRES | 4.12e-01 |

## ALIGNMENTS

RESULT 1  
ID ASPH\_HUMAN STANDARD; PRT; 757 AA.  
AC Q12797;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DE ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-  
DE HYDROXYLASE).  
DE DIOXYGENASE).  
GN ASPH  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 95121937.  
RA KORIOTH F., GIEFFERS C., FREY J.;  
RL GENE 150:395-399(1994).  
CC -!- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN  
CC CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF  
CC PROTEINS.  
CC -!- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) -  
CC PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).  
CC -!- COFACTOR: IRON.  
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC  
CC RETICULUM.  
CC -!- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES TESTED.  
CC -!- PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 274-757) OR 52 KD (AA  
CC 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY  
CC SIMILARITY).  
EMBL: U03109; E82591; -.  
DR MIM: 600582; -.  
DR OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR;  
KW ENDOPLASMIC RETICULUM.  
FT DOMAIN 1 54  
FT TRANSMEM 55 75  
FT DOMAIN 76 757  
FT DOMAIN 13 20  
FT DOMAIN 323 332  
FT CARBOHYD 452 452  
FT CARBOHYD 705 705  
FT SEQUENCE 757 AA; 85498 MW; AE6AFC24 CRC32;  
Query Match 100.0%; Score 41; DB 1; Length 757;  
Best Local Similarity 100.0%; Pred. No. 5.09e-01;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 175 PTGEPQ 180  
|||||  
QY 3 PTGEPQ 8

RESULT 2  
ID VG08\_HSV11 STANDARD; PRT; 232 AA.  
AC Q00137;  
DT 01-DEC-1992 (REL. 24, CREATED)  
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL GENE 8 MEMBRANE PROTEIN.  
GN 8.  
OS ICTALURID HERPESVIRUS 1 (CHANNEL CATFISH VIRUS) (CCV).  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AUBURN 1;  
RX MEDLINE; 92087490.  
RA DAVISON A.J.;  
RL VIROLOGY 186:9-14(1992).  
DR EMBL; M75136; G331295; -.  
DR EMBL; M75136; G331218; -.  
DR PIR; I36786; MMBE13.  
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.  
FT TRANSMEM 166 182 POTENTIAL.  
FT TRANSMEM 195 211 POTENTIAL.  
SQ SEQUENCE 232 AA; 25466 MW; AA90E31E CRC32;

Query Match 92.7%; Score 38; DB 1; Length 232;

Best Local Similarity 83.3%; Pred. No. 3.60e+00;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 90 PTGEPQ 95  
|||||  
QY 3 PTGEPQ 8

RESULT 3  
ID SYE\_RHIME STANDARD; PRT; 485 AA.  
AC P15189;  
DT 01-APR-1990 (REL. 14, CREATED)  
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
DE GLUTAMYL-TRNA SYNTHETASE (EC 6.1.1.17) (GLUTAMATE--TRNA LIGASE)  
DE (GLURS).  
GN GLTX.  
OS RHIZOBIUM MELILOTI.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
OC RHIZOBIACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A2;  
RX MEDLINE; 89291743.  
RA LABERGE S.; GAGNON Y.; BORDELEAU L.M.; LAPOINTE J.;  
RL J. BACTERIOL. 171:3926-3932(1989).  
CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + TRNA(GLU) = AMP +  
CC PYROPHOSPHATE + L-GLUTAMYL-TRNA(GLU).  
CC -!- SUBUNIT: MONOMER.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.  
DR EMBL; M27221; G717082; -.  
DR PIR; A32888; SYRZET.  
DR PROSITE; PS00178; AA-TRNA LIGASE I; 1.  
KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.  
FT SIMILAR 12 22 "HIGH" REGION.  
FT SIMILAR 253 257 "KMSK" REGION.  
FT BINDING 256 256 ATP (BY SIMILARITY).  
SQ SEQUENCE 485 AA; 54356 MW; 22131735 CRC32;

Query Match 92.7%; Score 38; DB 1; Length 485;

Best Local Similarity 83.3%; Pred. No. 3.60e+00;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 14 PTGEPH 19  
|||||  
QY 3 PTGEPQ 8

RESULT 4  
ID YRV2\_CAEEEL STANDARD; PRT; 518 AA.  
AC Q27519;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE PUTATIVE CYTOCHROME P450 T10B9.10 IN CHROMOSOME II (EC 1.14.-.-).  
GN T10B9.10.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA GARDNER A.;  
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED  
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
DR EMBL; Z48717; G732602; -.  
DR WORMPEP; T10B9.10; CE01655.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KW HYPOTHETICAL PROTEIN; OXIDOREDUCTASE; MONOOXYGENASE; HEME.  
FT BINDING 464 464 HEME (BY SIMILARITY).  
SQ SEQUENCE 518 AA; 58999 MW; F66B1048 CRC32;

Query Match 92.7%; Score 38; DB 1; Length 518;

Best Local Similarity 83.3%; Pred. No. 3.60e+00;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 270 PSGEPO 275  
|||||  
QY 3 PTGEPQ 8

RESULT 5  
ID YRV2\_CAEEEL STANDARD; PRT; 520 AA.  
AC Q27514;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE PUTATIVE CYTOCHROME P450 T10B9.2 IN CHROMOSOME II (EC 1.14.-.-).  
GN T10B9.2.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA GARDNER A.;  
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED  
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
DR EMBL; Z48717; G732596; -.  
DR WORMPEP; T10B9.2; CE01656.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KW HYPOTHETICAL PROTEIN; OXIDOREDUCTASE; MONOOXYGENASE; HEME.  
FT BINDING 464 464 HEME (BY SIMILARITY).  
SQ SEQUENCE 520 AA; 59524 MW; 087B5D9B CRC32;

Query Match 92.7%; Score 38; DB 1; Length 520;

Best Local Similarity 83.3%; Pred. No. 3.60e+00;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 270 PSCEPQ 275  
1:||||  
QY 3 PTGEPQ 8

RESULT 6  
ID XRV1\_CAEEL STANDARD; PRT; 520 AA.  
AC Q27513;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE PUTATIVE CYTOCHROME P450 T10B9.1 IN CHROMOSOME II (EC 1.14.-.-).  
GN T10B9.1.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA GARDNER A.;  
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
DR EMBL; Z48717; G732595; -.  
DR WORMPEP; T10B9.1; CE01654.  
DR PROSITE; PS00086; CYTOCHROME P450; 1.  
KW HYPOTHETICAL PROTEIN: OXIDOREDUCTASE; MONOOXYGENASE; HEME.  
FT BINDING 464 464 HEME (BY SIMILARITY).  
SQ SEQUENCE 520 AA; 59371 MW; 59371 MW; CFAVD0F8 CRC32;

Query Match 92.7%; Score 38; DB 1; Length 520;  
Best Local Similarity 83.3%; Pred. No. 3.60e+00;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 270 PSCEPQ 275  
1:||||  
QY 3 PTGEPQ 8

RESULT 7  
ID PT10\_STVPL STANDARD; PRT; 116 AA.  
AC P28202;  
DT 01-DEC-1992 (REL. 24, CREATED)  
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
DE PROTEIN-TYROSINE PHOSPHATASE 10 (EC 3.1.3.48) (FRAGMENT).  
GN STY 10.  
OS STYELA PLICATA (SEA SQUIRT).  
OC EUKARYOTA; METAZOA; CHORDATA; TUNICATA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 91139172.  
RA MATTHEWS R.J., FLORES E., THOMAS M.L.;  
RL IMMUNOGENETICS 33:33-41(1991).  
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O -> PROTEIN TYROSINE + ORTHOPHOSPHATE.  
CC EMBL; M37995; G160882; -.  
DR HSSP; P18031; 2HNP.  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; PARTIAL.  
DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; PARTIAL.  
DR PROSITE; PS00055; TYR\_PHOSPHATASE\_FTP; 1.  
KW HYDROLASE.  
FT NON\_TER 1  
FT NON\_TER 116  
SQ SEQUENCE 116 AA; 13251 MW; F7568322 CRC32;

Query Match 90.2%; Score 37; DB 1; Length 116;  
Best Local Similarity 66.7%; Pred. No. 6.75e+00;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 31 PSQDPQ 36  
1:||||

QY 3 PTGEPQ 8

RESULT 8  
ID VP6\_BT2A STANDARD; PRT; 301 AA.  
AC P33422;  
DT 01-FEB-1994 (REL. 28, CREATED)  
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
DE VP6 PROTEIN (MINOR INNER CORE PROTEIN VP6).  
GN S9.  
OS BLUETONGUE VIRUS (SEROTYPE 2 / ISOLATE USA).  
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 93033709.  
RA HWANG G.-Y., CHIOU J.-F., YANG Y.-Y., LI J.K.-K.;  
RL VIRUS RES. 24:315-323(1992).

CC -!- FUNCTION: SURROUNDS AND INTERACTS WITH THE GENOMIC DS-RNA. POSSESSES SS- AND DS-RNA-BINDING CAPACITY. ITS HYDROPHILIC NATURE AND CAPABILITY TO BIND SS- AND DS-RNA SUGGEST ITS INTERACTIONS WITH BTV GENOMIC RNA.  
CC -!- SUBCELLULAR LOCATION: INNER CAPSID.  
DR EMBL; L08668; G210851; -.  
KW CORE PROTEIN.  
FT DOMAIN 60 102 GLY-RICH.  
SQ SEQUENCE 301 AA; 32864 MW; B21871B4 CRC32;

Query Match 90.2%; Score 37; DB 1; Length 301;  
Best Local Similarity 66.7%; Pred. No. 6.75e+00;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 261 PTGDPH 266  
1:||||  
QY 3 PTGEPQ 8

RESULT 9  
ID VP6\_BT2A STANDARD; PRT; 325 AA.  
AC P32933;  
DT 01-OCT-1993 (REL. 27, CREATED)  
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
DE VP6 PROTEIN (MINOR INNER CORE PROTEIN VP6).  
GN S9.  
OS BLUETONGUE VIRUS (SEROTYPE 11 / ISOLATE USA).  
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 93033709.  
RA HWANG G.-Y., CHIOU J.-F., YANG Y.-Y., LI J.K.-K.;  
RL VIRUS RES. 24:315-323(1992).

CC -!- FUNCTION: SURROUNDS AND INTERACTS WITH THE GENOMIC DS-RNA. POSSESSES SS- AND DS-RNA-BINDING CAPACITY. ITS HYDROPHILIC NATURE AND CAPABILITY TO BIND SS- AND DS-RNA SUGGEST ITS INTERACTIONS WITH BTV GENOMIC RNA.  
CC -!- SUBCELLULAR LOCATION: INNER CAPSID.  
DR EMBL; L08670; G210845; -.  
DR PIR; A48561; A48561.  
KW CORE PROTEIN.  
FT DOMAIN 84 125 GLY-RICH.  
SQ SEQUENCE 325 AA; 35371 MW; 869B26B3 CRC32;

Query Match 90.2%; Score 37; DB 1; Length 325;  
Best Local Similarity 66.7%; Pred. No. 6.75e+00;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 285 PTGDPH 290  
1:||||  
QY 3 PTGEPQ 8

RESULT 10

ID VP6\_BT17 STANDARD; PRT; 325 AA.  
AC P32935;  
DT 01-OCT-1993 (REL. 27, CREATED)  
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
DE VP6 PROTEIN (MINOR INNER CORE PROTEIN VP6).  
GN S9.  
OS BLUETONGUE VIRUS (SEROTYPE 17 / ISOLATE USA).  
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 93033709.  
RA HWANG G.-Y., CHIOU J.-F., YANG Y.-Y., LI J.K.-K.;  
RL VIRUS RES. 24:315-323(1992).  
CC -1- FUNCTION: SURROUNDS AND INTERACTS WITH THE GENOMIC DS-RNA.  
CC POSSESSES SS- AND DS-RNA-BINDING CAPACITY. ITS HYDROPHILIC  
CC NATURE AND CAPABILITY TO BIND SS- AND DS-RNA SUGGEST ITS  
CC INTERACTIONS WITH BTV GENOMIC RNA.  
CC -1- SUBCELLULAR LOCATION: INNER CAPSID.  
DR EMBL; L08672; G210849; -.  
KW CORE PROTEIN.  
FT DOMAIN 84 125 GLY-RICH.  
SQ SEQUENCE 325 AA; 5D26438B CRC32;  
Query Match 90.2%; Score 37; DB 1; Length 325;  
Best Local Similarity 66.7%; Pred. No. 6.75e+00;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 285 PTGDPH 290  
|||:|  
QY 3 PTGEPQ 8  
RESULT 11  
ID VP6\_BT10 STANDARD; PRT; 325 AA.  
AC P33423;  
DT 01-FEB-1994 (REL. 28, CREATED)  
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
DE VP6 PROTEIN (MINOR INNER CORE PROTEIN VP6) (VERSION 2).  
GN S9.  
OS BLUETONGUE VIRUS (SEROTYPE 10 / ISOLATE USA).  
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 93033709.  
RA HWANG G.-Y., CHIOU J.-F., YANG Y.-Y., LI J.K.-K.;  
RL VIRUS RES. 24:315-323(1992).  
CC -1- FUNCTION: SURROUNDS AND INTERACTS WITH THE GENOMIC DS-RNA.  
CC POSSESSES SS- AND DS-RNA-BINDING CAPACITY. ITS HYDROPHILIC  
CC NATURE AND CAPABILITY TO BIND SS- AND DS-RNA SUGGEST ITS  
CC INTERACTIONS WITH BTV GENOMIC RNA.  
CC -1- CAUTION: SEE ALSO VERSION 1 OF THIS PROTEIN.  
DR EMBL; L08669; G210843; -.  
KW PIR; B48561; B48561.  
KW CORE PROTEIN.  
FT DOMAIN 85 125 GLY-RICH.  
SQ SEQUENCE 325 AA; 1141118E CRC32;  
Query Match 90.2%; Score 37; DB 1; Length 325;  
Best Local Similarity 66.7%; Pred. No. 6.75e+00;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 285 PTGDPH 290  
|||:|  
QY 3 PTGEPQ 8

RESULT 12  
ID VP6\_BT13 STANDARD; PRT; 325 AA.  
AC P32934;  
DT 01-OCT-1993 (REL. 27, CREATED)

DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
DE VP6 PROTEIN (MINOR INNER CORE PROTEIN VP6).  
GN S9.  
OS BLUETONGUE VIRUS (SEROTYPE 13 / ISOLATE USA).  
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 93033709.  
RA HWANG G.-Y., CHIOU J.-F., YANG Y.-Y., LI J.K.-K.;  
RL VIRUS RES. 24:315-323(1992).  
CC -1- FUNCTION: SURROUNDS AND INTERACTS WITH THE GENOMIC DS-RNA.  
CC POSSESSES SS- AND DS-RNA-BINDING CAPACITY. ITS HYDROPHILIC  
CC NATURE AND CAPABILITY TO BIND SS- AND DS-RNA SUGGEST ITS  
CC INTERACTIONS WITH BTV GENOMIC RNA.  
CC -1- SUBCELLULAR LOCATION: INNER CAPSID.  
DR EMBL; L08671; G210847; -.  
KW CORE PROTEIN.  
FT DOMAIN 84 125 GLY-RICH.  
SQ SEQUENCE 325 AA; 7F5B6F92 CRC32;  
Query Match 90.2%; Score 37; DB 1; Length 325;  
Best Local Similarity 66.7%; Pred. No. 6.75e+00;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 285 PTGDPH 290  
|||:|  
QY 3 PTGEPQ 8  
RESULT 13  
ID VP6\_BT15 STANDARD; PRT; 328 AA.  
AC Q03328; Q65752;  
DT 01-OCT-1993 (REL. 27, CREATED)  
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE VP6 PROTEIN (MINOR INNER CORE PROTEIN VP6).  
GN S9.  
OS BLUETONGUE VIRUS (SEROTYPE 1 / ISOLATE SOUTH AFRICA).  
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 93057380.  
RA WADE-EVANS A.M., MERTENS P.P.C., BELSHAM G.J.;  
RL J. GEN. VIROL. 73:3023-3026(1992).  
CC -1- FUNCTION: SURROUNDS AND INTERACTS WITH THE GENOMIC DS-RNA.  
CC POSSESSES SS- AND DS-RNA-BINDING CAPACITY. ITS HYDROPHILIC  
CC NATURE AND CAPABILITY TO BIND SS- AND DS-RNA SUGGEST ITS  
CC INTERACTIONS WITH BTV GENOMIC RNA.  
CC -1- SUBCELLULAR LOCATION: INNER CAPSID.  
DR EMBL; D10905; G221081; -.  
DR PIR; JQ1875; JQ1875.  
KW CORE PROTEIN.  
FT CHAIN 1 328 VP6.  
FT CHAIN 5 328 VP6A.  
FT DOMAIN 88 130 GLY-RICH.  
SQ SEQUENCE 328 AA; 35928 MW; D48A20F2 CRC32;  
Query Match 90.2%; Score 37; DB 1; Length 328;  
Best Local Similarity 66.7%; Pred. No. 6.75e+00;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 289 PTGDPH 294  
|||:|  
QY 3 PTGEPQ 8  
RESULT 14  
ID VP61\_BT10 STANDARD; PRT; 329 AA.  
AC P23066;  
DT 01-NOV-1991 (REL. 20, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)

DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE VP6 PROTEIN (MINOR INNER CORE PROTEIN VP6) (VERSION 1).  
GN S9.  
OS BLUTETONGUE VIRUS (SEROTYPE 10 / ISOLATE USA).  
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 89293076.  
RA FUKUSHO A., YU Y., YAMAGUCHI S., ROY P.;  
RL J. GEN. VIROL. 70:1677-1689(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90345726.  
RA ROY P., MARSHALL J.J.A., FRENCH T.J.;  
RL CURR. TOP. MICROBIOL. IMMUNOL. 162:43-87(1990).  
RN [3]  
RP REVISIONS.

RA BLACHERE C.;  
RL SUBMITTED (JAN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- FUNCTION: SURROUNDS AND INTERACTS WITH THE GENOMIC DS-RNA.  
CC POSSESSES SS- AND DS-RNA-BINDING CAPACITY. ITS HYDROPHILIC  
CC NATURE AND CAPABILITY TO BIND SS- AND DS-RNA SUGGEST ITS  
CC INTERACTIONS WITH BTV GENOMIC RNA.  
CC -!- SUBCELLULAR LOCATION: INNER CAPSID.  
CC -!- CAUTION: SEE ALSO VERSION 2 OF THIS PROTEIN.  
DR EMBL; D00509; G221077; -.  
DR PIR; B32400; VEXRG3.  
DR PIR; S10542; S10542.  
KW CORE PROTEIN.88 130 GLY-RICH.  
FT DOMAIN

FT SEQUENCE 329 AA; 35515 MW; 701CD715 CRC32;

Query Match 90.2%; Score 37; DB 1; Length 370;  
Best Local Similarity 66.7%; Pred. No. 6.75e+00;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 16 PTGDPH 21  
QY 3 PTGEPQ 8

Search completed: Thu May 13 15:41:13 1999  
Job time : 8 secs.

Query Match 90.2%; Score 37; DB 1; Length 329;  
Best Local Similarity 66.7%; Pred. No. 6.75e+00;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294  
QY 3 PTGEPQ 8

RESULT 15  
ID SVE\_CHLPS STANDARD; PRT; 370 AA.  
AC Q06560;  
DT 01-JUN-1994 (REL. 29, CREATED)  
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
DE GLUTAMYL-TRNA SYNTHETASE (EC 6.1.1.17) (GLUTAMATE--TRNA LIGASE)  
DE (GLURS) (FRAGMENT).  
GN GLTX.  
OS CHLAMYDIA PSITTACI.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; RICKETTSIAS; CHLAMYDIALES;  
OC CHLAMYDIACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=6BC;  
RX MEDLINE; 93259937.  
RA WICHIAN D.W., HATCH T.P.;  
RL J. BACTERIOL. 175:2936-2942(1993).  
CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + TRNA(GLU) = AMP +  
CC PYROPHOSPHATE + L-GLUTAMYL-TRNA(GLU).  
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.  
DR EMBL; L13598; G289833; -.  
DR PIR; A36909; A36909.  
DR PROSITE; PS00178; AA-TRNA-LIGASE\_I; 1.  
KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.  
FT SIMILAR 14 24 "HIGH" REGION.  
FT SIMILAR 255 259 "RMSKS" REGION.  
FT BINDING 258 258 ATP (BY SIMILARITY).  
FT NON-TER 370 370  
FT SEQUENCE 370 AA; 43093 MW; 5F53F41B CRC32;

\*\*\*\*\*  
M P E R L A  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu May 13 15:44:14 1999; MasPar time 2.54 Seconds  
Tabular output not generated. 63.620 Million cell updates/sec

Title: >US-09-040-485-9  
Description: (1-10) from US09040485.pep  
Perfect Score: 61  
Sequence: 1 QENPDSSEPV 10

Scoring table: PAM 150  
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq32  
1:part1 2:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 14.163; Variance 44.856; scale 0.316

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID        | Description           | Pred. No. |
|------------|-------|-------------|--------|-----------|-----------------------|-----------|
| 1          | 43    | 70.5        | 995    | 12 R60812 | Agarase 0107 from Vib | 1.67e+02  |
| 2          | 42    | 68.9        | 332    | 20 W06491 | Beta-1-4-galactosyltr | 2.13e+02  |
| 3          | 41    | 67.2        | 14     | 29 W40864 | Cytotoxic Epstein-Bar | 2.71e+02  |
| 4          | 41    | 67.2        | 84     | 13 R66800 | N-terminal fragment ( | 2.71e+02  |
| 5          | 41    | 67.2        | 154    | 13 R66801 | N-terminal fragment ( | 2.71e+02  |
| 6          | 41    | 67.2        | 188    | 13 R66802 | N-terminal fragment ( | 2.71e+02  |
| 7          | 41    | 67.2        | 228    | 13 R66804 | N-terminal fragment ( | 2.71e+02  |
| 8          | 41    | 67.2        | 289    | 13 R66796 | Novel mouse proteogly | 2.71e+02  |
| 9          | 41    | 67.2        | 311    | 29 W47156 | Mouse syndecan protei | 2.71e+02  |
| 10         | 41    | 67.2        | 311    | 13 R66793 | Novel mouse proteogly | 2.71e+02  |
| 11         | 41    | 67.2        | 311    | 10 R55276 | Syndecan protein.     | 2.71e+02  |
| 12         | 41    | 67.2        | 311    | 17 R87001 | Mouse syndecan-1.     | 2.71e+02  |
| 13         | 41    | 67.2        | 919    | 23 W15850 | Potato alpha-glucosid | 2.71e+02  |
| 14         | 40    | 65.6        | 10     | 27 W42736 | Antigenic decapeptide | 3.43e+02  |
| 15         | 40    | 65.6        | 10     | 28 W42824 | Antigenic decapeptide | 3.43e+02  |
| 16         | 40    | 65.6        | 151    | 27 W42642 | Human sperm zona bind | 3.43e+02  |
| 17         | 40    | 65.6        | 151    | 14 R79761 | Human testis sperm zo | 3.43e+02  |
| 18         | 40    | 65.6        | 162    | 28 W42692 | Recombinant human Spi | 3.43e+02  |

|    |    |      |      |           |                        |          |
|----|----|------|------|-----------|------------------------|----------|
| 19 | 40 | 65.6 | 163  | 28 W42693 | Baboon sperm zona bin  | 3.43e+02 |
| 20 | 40 | 65.6 | 331  | 26 W29817 | Mammalian AMPK-gamma   | 3.43e+02 |
| 21 | 40 | 65.6 | 461  | 27 W33890 | Flea saliva protein P  | 3.43e+02 |
| 22 | 39 | 63.9 | 226  | 3 R13901  | Nitrile hydratase L t  | 4.34e+02 |
| 23 | 39 | 63.9 | 260  | 21 W14422 | Human Dnase I variant  | 4.34e+02 |
| 24 | 39 | 63.9 | 496  | 22 W08605 | Chimeric MOMLV and Ty  | 4.34e+02 |
| 25 | 39 | 63.9 | 517  | 22 W08607 | Chimeric MOMLV and Ty  | 4.34e+02 |
| 26 | 39 | 63.9 | 866  | 19 W04185 | Human Interleukin-17   | 4.34e+02 |
| 27 | 39 | 63.9 | 2019 | 13 R67913 | Cardiac sodium channe  | 4.34e+02 |
| 28 | 39 | 63.9 | 2020 | 2 R6584   | Cardiac sodium channe  | 4.34e+02 |
| 29 | 38 | 62.3 | 147  | 27 W36001 | Human Fhit protein.    | 5.48e+02 |
| 30 | 38 | 62.3 | 147  | 28 W39124 | Human PSAFP-1 protein  | 5.48e+02 |
| 31 | 38 | 62.3 | 537  | 29 W46736 | Amino acid sequence o  | 5.48e+02 |
| 32 | 38 | 62.3 | 591  | 29 W41418 | Paxillin beta isoform  | 5.48e+02 |
| 33 | 38 | 62.3 | 605  | 29 W41419 | Paxillin gamma isoform | 5.48e+02 |
| 34 | 38 | 62.3 | 1100 | 1 P95644  | Rabbit skeletal muscle | 5.48e+02 |
| 35 | 38 | 62.3 | 1103 | 14 R71012 | Human neuronal calciu  | 5.48e+02 |
| 36 | 38 | 62.3 | 1106 | 22 W18389 | Rabbit calcium channe  | 5.48e+02 |
| 37 | 38 | 62.3 | 1106 | 13 R73056 | Rabbit skeletal calci  | 5.48e+02 |
| 38 | 38 | 62.3 | 1106 | 26 W37712 | Rabbit skeletal calci  | 5.48e+02 |
| 39 | 38 | 62.3 | 1203 | 8 R41668  | Human endothelial cel  | 5.48e+02 |
| 40 | 38 | 62.3 | 1443 | 29 W44137 | Homo sapiens factor V  | 5.48e+02 |
| 41 | 38 | 62.3 | 2133 | 29 W44133 | Sus scrofa factor VII  | 5.48e+02 |
| 42 | 38 | 62.3 | 2986 | 8 R44432  | eryA region polypepti  | 5.48e+02 |
| 43 | 37 | 60.7 | 234  | 3 P50139  | Rabbit tumor necrosis  | 6.92e+02 |
| 44 | 37 | 60.7 | 552  | 3 P60956  | Sequence of bovine me  | 6.92e+02 |
| 45 | 37 | 60.7 | 761  | 9 R48036  | Mycobacterium BCG imm  | 6.92e+02 |

ALIGNMENTS

RESULT 1  
ID R60812 standard; Protein; 995 AA.

AC R60812;

DT 20-JUL-1995 (first entry)

DE Agarase 0107 from Vibrio species.

KW Agarase; Vibrio; PCR; amplification; amplify; primer; probe; E.coli;

OS cation-exchange; chromatograph; neogaroooligosaccharide.

FS Vibrio sp. J70107.

FT Key Location/Qualifiers

FT peptide 1..20

FT protein /label= signal peptide

FT protein 21..995

FT protein /label= mature peptide

PN J0628488-A.

PD 11-OCT-1994.

PF 01-APR-1993; 096549.

PR 01-APR-1993; JP-096549

PA (NIBS ) JAPAN TOBACCO INC.

DR WPI; 94-362595/45.

DR N-PSDB; Q73755.

PT DNA encoding protein with agarase activity - useful for the

PT industrial production of a large amount of

PT neo-agarao-oligosaccharide

PS Claim 1; Page 8-12; 13pp; Japanese.

CC The sequence of a novel protein containing agarase activity from

CC Vibrio species. The gene of 3205 bp encodes a protein of 995 amino

CC acids. The gene was obtained by PCR amplification from a Vibrio genomic

CC library using primers and probes based on the amino acid sequence

CC (R60813-21) from protein sequencing of partially degraded agarase 0107.

CC The gene was then introduced into E.coli JM83. The expressed protein

CC could then be recovered by cation-exchange chromatography. The protein

CC can be used in the large scale manufacture of neogaroooligosaccharide.

SQ Sequence 995 AA;

Query Match 70.5%; Score 43; DB 12; Length 995;

Best Local Similarity 66.7%; Pred. No. 1.67e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 947 eapdpgepv 955

Qy 2 ENPDSSEPV 10

```

RESULT 2
ID W06491 standard; Protein; 332 AA.
AC W06491;
DE 03-FEB-1997 (first entry)
DT Beta-1,4-galactosyltransferase-related protein #2.
KW Murine; beta-1,4-galactosyltransferase-related protein; sterility;
KW fertilisation; F9 cancer cell; Huynh's method.
OS Mus musculus.
PN J08196279-A.
PD 06-AUG-1996.
PF 25-JAN-1995; 009642.
PR 25-JAN-1995; JP-009642.
PA (MITK ) MITSUI TOATSU CHEM INC.
PA (MURA) MORAWATSU T.
DR WPI; 96-406013/41.
DR N-PSDB; T45082.
PT DNA sequence encoding beta-1,4-galactosyltransferase-related
PT protein - useful for sterility diagnosis, and for assisting or
PT inhibiting fertilisation
PS Claim 4; Page 7-9; ilpp; Japanese.
CC The sequences given in W06490-91 represent two clones of murine
CC beta-1,4-galactosyltransferase-related proteins. These proteins
CC can be used as diagnostic agents for various diseases. They are
CC esp. useful in the diagnosis of sterility and in the aiding and
CC inhibiting of fertilisation. The cDNA's encoding the two beta-1,4-
CC galactosyltransferase-related proteins were isolated from F9 cancer
CC cells according to Huynh's method.
SQ Sequence 332 AA;

Query Match 68.9%; Score 42; DB 20; Length 332;
Best Local Similarity 55.6%; Pred. No. 2.13e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 226 dspdsdvpv 234
QY :|||:|
2 ENPDSSEPV 10

RESULT 3
ID W40864 standard; peptide; 14 AA.
AC W40864;
DE 09-JUN-1998 (first entry)
DT Cytotoxic Epstein-Barr T-cell epitope 37.
KW Herpes simplex virus type 4; Epstein-Barr virus; EBV; cytotoxic;
KW T-cell epitope; nuclear antigen; human leukocyte antigen; HLA; vaccine;
KW tetanus toxoid; diphtheria toxoid; Bordetella pertussis;
KW poliovirus antigen.
OS Herpes simplex virus type 4.
PN W09745444-A1.
PD 04-DEC-1997.
PF 23-MAY-1997; AU0328.
PR 24-MAY-1996; AU-000073.
PA (BIOT-) BIOTECH AUSTRALIA PTY LTD.
PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA (COUN-) COUNCIL QUEENSLAND INST MEDICAL RES.
PA (CSLC-) CSL LTD.
PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
PA (UYME ) UNIV MELBOURNE.
PI Burrows SR, Kerr BM, Khanna R, Misko IS, Moss DJ;
DR WPI; 98-032576/03.
PT Cytotoxic Epstein-Barr virus T cell epitope - useful to prepare
PT vaccine to prevent and treat infection
PS Claim 3; Page 32; 41pp; English.
CC This amino acid sequence is a cytotoxic Epstein-Barr virus (EBV) T-cell
CC epitope of the EBV nuclear antigen 3A (EBNA3A), which binds the human
CC leukocyte antigen acceptor A11 (HLA A11). It is used to prepare a
CC vaccine which may include the cytotoxic EBV T-cell epitope (W40828-
CC W40846), or a nucleic acid sequence encoding it. The vaccine produced
CC comprises or encodes at least one antigen (W40847-W40876) to which the
CC individual will mount an anamnestic response, e.g. a tetanus toxoid,
CC diphtheria toxoid, Bordetella pertussis antigen, poliovirus antigen,
CC purified protein derivative, glycoprotein 350 protein, and/or helper

```

---

```

CC epitope.
SQ Sequence 14 AA;

Query Match 67.2%; Score 41; DB 29; Length 14;
Best Local Similarity 55.6%; Pred. No. 2.71e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 1 depasteqv 9
QY :|||:|
2 ENPDSSEPV 10

RESULT 4
ID R66800 standard; Peptide; 84 AA.
AC R66800;
DE 11-SEP-1995 (first entry)
DT N-terminal fragment (a.a. 1-84) of mature mouse syndecan-1.
KW Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;
KW ectodomain; NMuMG; glycosylation; heparan sulphate; chondroitin sulphate;
KW glycosaminoglycan; chimaera; chimaeric molecule; effector molecule;
KW receptor; drug; antibody; diagnostic agent.
OS Mus musculus.
PN W09500633-A.
PD 05-JAN-1995.
PF 17-JUN-1994; U06920.
PR 17-JUN-1993; US-078683.
PA (CHIL-) CHILDRENS MEDICAL CENT.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Bernfield M, Kato M, Saunders S;
DR WPI; 95-052071/07.
PT DNA and protein sequences for recombinant syndecan-derived
PT proteoglycans - comprising a core protein having glycosylation
PT sites for heparin sulphate glycosaminoglycan side chains.
PS Claim 15; Page 82; 97pp; English.
CC Peptides R66797-802 correspond to varying length fragments of the
CC N-terminal sequence of the mature mouse syndecan-1 (see R66793 and
CC R66796). The sequence given here covers the amino acids from 1-84 of the
CC mature protein sequence. The fragments are portions of the extracellular
CC domain of the protein and contain a heparan sulphate glycosaminoglycan
CC attachment site (R66795). The functional domains, esp. the soluble
CC extracellular or heparan binding site, of the syndecan molecules (see
CC R66797-812 and R66818) can be used to construct chimaeras by linking them
CC to biological effector molecules, cell surface receptors, drugs,
CC antibodies, diagnostic agents or components of microorganisms.
SQ Sequence 84 AA;

Query Match 67.2%; Score 41; DB 13; Length 84;
Best Local Similarity 55.6%; Pred. No. 2.71e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 74 ekpeegepv 82
QY :|||:|
2 ENPDSSEPV 10

RESULT 5
ID R66801 standard; Peptide; 154 AA.
AC R66801;
DE 11-SEP-1995 (first entry)
DT N-terminal fragment (a.a. 1-154) of mature mouse syndecan-1.
KW Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;
KW ectodomain; NMuMG; glycosylation; heparan sulphate; chondroitin sulphate;
KW glycosaminoglycan; chimaera; chimaeric molecule; effector molecule;
KW receptor; drug; antibody; diagnostic agent.
OS Mus musculus.
PN W09500633-A.
PD 05-JAN-1995.
PF 17-JUN-1994; U06920.
PR 17-JUN-1993; US-078683.
PA (CHIL-) CHILDRENS MEDICAL CENT.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Bernfield M, Kato M, Saunders S;
DR WPI; 95-052071/07.

```

PT DNA and protein sequences for recombinant syndecan-derived  
 PT proteoglycans - comprising a core protein having glycosylation  
 PT sites for heparin sulphate glycosaminoglycan side chains.  
 PS Claim 15; Page 82; 97pp; English.  
 CC Peptides R66797-802 correspond to varying length fragments of the  
 CC N-terminal sequence of the mature mouse syndecan-1 (see R66793 and  
 CC R66796). The sequence given here covers the amino acids from 1-154 of the  
 CC mature protein sequence. The fragments are portions of the extracellular  
 CC domain of the protein and contain a heparan sulphate glycosaminoglycan  
 CC attachment site (R66795). The functional domains, esp. the soluble  
 CC extracellular or heparan binding site, of the syndecan molecules (see  
 CC R66797-812 and R66818) can be used to construct chimaeras by linking them  
 CC to biological effector molecules, cell surface receptors, drugs,  
 CC antibodies, diagnostic agents or components of microorganisms.  
 SQ Sequence 154 AA;

Query Match 67.2%; Score 41; DB 13; Length 154;  
 Best Local Similarity 55.6%; Pred. No. 2.71e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 74 ekpeegpv 82  
 I::: :|||  
 QY 2 ENPDSSEPV 10

## RESULT 6

ID R66802 standard; Peptide; 188 AA.  
 AC R66802;  
 DT 11-SEP-1995 (first entry)  
 DE N-terminal fragment (a.a. 1-188) of mature mouse syndecan-1.  
 KW Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;  
 KW ectodomain; NMuMG; glycosylation; heparan sulphate; chondroitin sulphate;  
 KW glycosaminoglycan; chimaera; chimaeric molecule; effector molecule;  
 KW receptor; drug; antibody; diagnostic agent.  
 OS Mus musculus.  
 PN W09500633-A.  
 PD 05-JAN-1995.  
 PF 17-JUN-1994; U06920.  
 PR 17-JUN-1993; US-078683.  
 PA (CHIL-) CHILDRENS MEDICAL CENT.  
 PA (STRD) UNIV LELAND STANFORD JUNIOR.  
 PI Bernfield M, Kato M, Saunders S;  
 DR WPI; 95-052071/07.  
 PT DNA and protein sequences for recombinant syndecan-derived  
 PT proteoglycans - comprising a core protein having glycosylation  
 PT sites for heparin sulphate glycosaminoglycan side chains.  
 PS Claim 15; Page 82; 97pp; English.  
 CC Peptides R66797-802 correspond to varying length fragments of the  
 CC N-terminal sequence of the mature mouse syndecan-1 (see R66793 and  
 CC R66796). The sequence given here covers the amino acids from 1-188 of the  
 CC mature protein sequence. The fragments are portions of the extracellular  
 CC domain of the protein and contain a heparan sulphate glycosaminoglycan  
 CC attachment site (R66795). The functional domains, esp. the soluble  
 CC extracellular or heparan binding site, of the syndecan molecules (see  
 CC R66797-812 and R66818) can be used to construct chimaeras by linking them  
 CC to biological effector molecules, cell surface receptors, drugs,  
 CC antibodies, diagnostic agents or components of microorganisms.  
 SQ Sequence 188 AA;

Query Match 67.2%; Score 41; DB 13; Length 188;  
 Best Local Similarity 55.6%; Pred. No. 2.71e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 74 ekpeegpv 82  
 I::: :|||  
 QY 2 ENPDSSEPV 10

## RESULT 7

ID R66804 standard; Peptide; 228 AA.  
 AC R66804;  
 DT 11-SEP-1995 (first entry)  
 DE N-terminal fragment (a.a. 1-228) of mature mouse syndecan-1.

KW Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;  
 KW ectodomain; NMuMG; glycosylation; heparan sulphate; chondroitin sulphate;  
 KW glycosaminoglycan; chimaera; chimaeric molecule; effector molecule;  
 KW receptor; drug; antibody; diagnostic agent.  
 OS Mus musculus.  
 PN W09500633-A.  
 PD 05-JAN-1995.  
 PF 17-JUN-1994; U06920.  
 PR 17-JUN-1993; US-078683.  
 PA (CHIL-) CHILDRENS MEDICAL CENT.  
 PA (STRD) UNIV LELAND STANFORD JUNIOR.  
 PI Bernfield M, Kato M, Saunders S;  
 DR WPI; 95-052071/07.  
 PT DNA and protein sequences for recombinant syndecan-derived  
 PT proteoglycans - comprising a core protein having glycosylation  
 PT sites for heparin sulphate glycosaminoglycan side chains.  
 PS Claim 22; Page 83; 97pp; English.  
 CC Peptides R66803-810 correspond to the extracellular domains of the mouse  
 CC syndecans 1-4 or fragments of these domains. The sequence given here  
 CC represents amino acids 1-228, thus covering the entire extracellular  
 CC domain, of the mature mouse syndecan-1 protein (see R66793 and R66796).  
 CC The peptides presented all contain a heparan sulphate glycosaminoglycan  
 CC attachment site (see R66794). The functional domains, esp. the soluble  
 CC extracellular or heparan binding site, of the syndecan molecules (see  
 CC R66797-812 and R66818) can be used to construct chimaeras by linking them  
 CC to biological effector molecules, cell surface receptors, drugs,  
 CC antibodies, diagnostic agents or components of microorganisms.  
 SQ Sequence 228 AA;

Query Match 67.2%; Score 41; DB 13; Length 228;  
 Best Local Similarity 55.6%; Pred. No. 2.71e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 74 ekpeegpv 82  
 I::: :|||  
 QY 2 ENPDSSEPV 10

## RESULT 8

ID R66796 standard; Protein; 289 AA.  
 AC R66796;  
 DT 11-SEP-1995 (first entry)  
 DE Novel mouse proteoglycan syndecan-1 putative mature protein sequence.  
 KW Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;  
 KW ectodomain; NMuMG; glycosylation; heparan sulphate; chondroitin sulphate;  
 KW chimaera; chimaeric molecule; effector molecule; receptor; drug;  
 KW antibody; diagnostic agent.  
 OS Mus musculus.  
 FH Key Location/Qualifiers  
 FT domain 1..230  
 FT domain /label= extracellular domain  
 FT domain 23..26  
 FT domain /note= "heparan sulphate attachment site"  
 FT domain 231..255  
 FT domain /label= transmembrane domain  
 FT domain 256..289  
 FT domain /label= intracellular domain  
 PN W09500633-A.  
 PD 05-JAN-1995.  
 PF 17-JUN-1994; U06920.  
 PR 17-JUN-1993; US-078683.  
 PA (CHIL-) CHILDRENS MEDICAL CENT.  
 PA (STRD) UNIV LELAND STANFORD JUNIOR.  
 PI Bernfield M, Kato M, Saunders S;  
 DR WPI; 95-052071/07.  
 PT DNA and protein sequences for recombinant syndecan-derived  
 PT proteoglycans - comprising a core protein having glycosylation  
 PT sites for heparin sulphate glycosaminoglycan side chains.  
 PS Claim 13; Page 81; 97pp; English.  
 CC The sequence of the mature cell surface proteoglycan - syndecan-1.  
 CC The corresponding gene (Q81748) was isolated from a mouse mammary  
 CC epithelial cell cDNA library in lambda gt10 using a rabbit antisera against  
 CC the ectodomain of NMuMG mouse mammary epithelial cell syndecan-1. The



CC complete sequence of the gene encodes a protein(R66793) of 30-35 kD  
 CC comprising: (1) a hydrophilic N-terminal extracellular domain (residues  
 CC 23-252); (2) a hydrophobic transmembrane domain (residues 253-277) and  
 CC (3) a hydrophilic C-terminal intracellular domain (residues 278-331).  
 CC The protein contains a protease susceptible cleavage sequence.  
 CC Extracellularly and adjacent to the transmembrane region and at least one  
 CC glycosylation site (residues 45-48) for attachment of a heparan sulphate  
 CC chain to the extracellular region. The syndecan-1 protein is thought to  
 CC contain a 22 amino acid signal peptide, lacking in this sequence, but  
 CC this was unresolved due to N-terminal blocking of the mature peptide,  
 CC however a likely site for signal peptidase cleavage occurs after Pro-22.  
 CC The functional domains, esp. the soluble extracellular or heparan binding  
 CC site, of the syndecan molecules (see R66797-812 and R66818) can be used  
 CC to construct chimaeras by linking them to biological effector molecules,  
 CC cell surface receptors, drugs, antibodies, diagnostic agents or  
 CC components of microorganisms.  
 CC Sequence 289 AA;

Query Match 67.2%; Score 41; DB 13; Length 289;  
 Best Local Similarity 55.6%; Pred. No. 2.71e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 74 ekpegepv 82  
 I::: |||  
 QY 2 ENPDSSEPV 10

RESULT 9  
 ID W47156 standard; Protein; 311 AA.

AC W47156;  
 DT 28-MAY-1998 (first entry)  
 DE Mouse syndecan protein sequence.  
 KW Syndecan; tumour suppression; tissue regeneration; enhancement;  
 KW mouse; wound healing.

OS Mus sp.

PN US5726058-A.

PD 10-MAR-1998.

PF 07-JUN-1995; 472217.

PR 07-MAR-1994; US-206186.

PR 01-DEC-1992; US-988427.

PR 01-DEC-1993; WO-FI0514.

PR 07-JUN-1995; US-472217.

PA (ALAN)/ ALANEN-KURKI L.

PA (AUVI)/ AUVINEN P.

PA (JAAK)/ JAAKKOLA P.

PA (JALK)/ JALKANEN M.

PA (LEPP)/ LEPPAE S.

PA (MALI)/ MALI M.

PA (VIHI)/ VIHINEN T.

PA (WAER)/ WAERRI A.

PI Alanen-kurki L, Auvinen P, Jaakkola P, Jalkanen M,

PI Leppae S, Mali M, Viuhinen T, Waerri A;

DR WPI: 98-192770/17.

DR N-PSDB: V15946.

PT New mouse syndecan gene sequences - useful for, e.g. suppressing  
 PT tumour growth or promoting tissue regeneration in processes such as  
 PT wound healing

PS Disclosure; Fig 2A-O: 48pp; English.

CC This is the mouse syndecan protein sequence. A 350 base pair fragment  
 CC (V15948) of a purified 2196 base pair DNA molecule (V15947) enhances  
 CC the expression of a gene operably linked to the promoter of the mouse  
 CC syndecan gene in 3f3 cells following treatment with rGF- beta and bFGF  
 CC when the fragment is operably linked to the promoter. A purified DNA  
 CC molecule comprising a portion of the nucleotide residues 3538-3888 of  
 CC the mouse syndecan genomic sequence suppresses expression of a gene  
 CC operably linked to the promoter of the mouse syndecan gene in 5f15 cells  
 CC treated with testosterone. Host cells can be transfected with vectors  
 CC which contain either the enhancing or suppressing DNA molecules. The  
 CC products may be used to alter the differentiated state of a host cell by  
 CC altering its expression of syndecan, e.g. to induce and regulate syndecan  
 CC expression, especially in cells which exhibit a malignant phenotype,  
 CC regardless of the origin of transformation. The products can be used to  
 CC produce therapeutics for suppressing tumour growth. They may enhance the

CC syndecan expression in a host cell, by enhancing its gene transcription,  
 CC especially in malignant or normal cells, and therefore promote tissue  
 CC regeneration, especially in processes such as wound healing.  
 SQ Sequence 311 AA;

Query Match 67.2%; Score 41; DB 29; Length 311;  
 Best Local Similarity 55.6%; Pred. No. 2.71e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 96 ekpegepv 104  
 I::: |||  
 QY 2 ENPDSSEPV 10

RESULT 10

ID R66793 standard; Protein; 311 AA.

AC R66793;

DT 11-SEP-1995 (first entry)

DE Novel mouse proteoglycan syndecan-1 protein sequence.

KW Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;

KW ectodomain; NMuMG; glycosylation; heparan sulphate; chondroitin sulphate;

KW chimaera; chimaeric molecule; effector molecule; receptor; drug;

KW antibody; diagnostic agent.

OS Mus musculus.

FS Key Location/Qualifiers

FT peptide 1..22

FT /label= putative signal peptide

FT domain 23..252

FT /label= extracellular domain

FT domain 45..48

FT /label= heparan sulphate attachment site

FT domain 253..277

FT /label= transmembrane domain

FT domain 278..311

FT /label= intracellular domain

PN WO9500633-A.

PD 05-JAN-1995.

PF 17-JUN-1994; U06920.

PR 17-JUN-1993; US-078683.

PA (CHIL-) CHILDRENS MEDICAL CENT.

PA (STRD) UNIV LELAND STANFORD JUNIOR.

PI Bernfield M, Kato M, Saunders S;

DR WPI: 95-052071/07.

DR N-PSDB: Q81748.

PT DNA and protein sequences for recombinant syndecan-derived

PT proteoglycans - comprising a core protein having glycosylation

PT sites for heparin sulphate glycosaminoglycan side chains.

PS Claim 13; Page 80: 97pp; English.

CC The sequence of the novel cell surface proteoglycan - syndecan-1. The  
 CC corresponding gene was isolated from a mouse mammary epithelial cell cDNA  
 CC library in lambda gt11 using rabbit antisera against the ectodomain of  
 CC NMuMG mouse mammary epithelial cell syndecan-1. The complete sequence of  
 CC the gene encodes a protein of 30-35 kD comprising: (1) a hydrophilic  
 CC N-terminal extracellular domain (residues 23-252); (2) a hydrophobic  
 CC transmembrane domain (residues 253-277) and (3) a hydrophilic C-terminal  
 CC intracellular domain (residues 278-331). The protein contains a protease  
 CC susceptible cleavage sequence extracellularly and adjacent to the  
 CC transmembrane region and at least one glycosylation site (residues 45-48)  
 CC for attachment of a heparan sulphate chain to the extracellular region.  
 CC The syndecan-1 protein is thought to contain a 22 amino acid signal  
 CC peptide but this was unresolved due to N-terminal blocking of the mature  
 CC peptide, however a likely site for signal peptidase cleavage occurs after  
 CC Pro-22. The functional domains, esp. the soluble extracellular or  
 CC heparan binding site, of the syndecan molecules (see R66797-812 and  
 CC R66818) can be used to construct chimaeras by linking them to biological  
 CC effector molecules, cell surface receptors, drugs, antibodies, diagnostic  
 CC agents or components of microorganisms.  
 SQ Sequence 311 AA;

Query Match 67.2%; Score 41; DB 13; Length 311;

Best Local Similarity 55.6%; Pred. No. 2.71e+02;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 96 ekpeesepv 104  
 QY 2 ENPDSSEPV 10

RESULT 11  
 ID R55276 standard; Protein; 311 AA.  
 AC R55276;  
 DT 08-DEC-1994 (first entry)  
 DE Syndecan protein.  
 KW Syndecan; enhancer; differentiation; tumor; therapy; hair growth.  
 OS Mus sp.  
 PN WO9412162-A.  
 PD 09-JUN-1994.  
 PF 01-DEC-1993; F10514.  
 PR 01-DEC-1992; US-988427.  
 PA (WAER/) WAERRI A M.  
 PA (ALAN/) ALANEN-KURKI L M.  
 PA (AUVI/) AUVINEN P O V.  
 PI Alanen-Kurki LM, Auvinen POV, Jaakkola PM, Jalkanen MT;  
 PI Leppaesm, Mali MS, Viuhinen TA, Waerri AM;  
 DR WPI: 94-199926/24.  
 DR N-PSDB: Q67902.  
 DT Syndecan stimulation of cellular differentiation - useful for  
 PT decreasing tumour growth used to promote hair growth  
 PS Disclosure; Page 22-39; 65pp; English.  
 CC The mouse syndecan gene enhancer, located 8-10 kb upstream from the  
 CC initiation site, is given in Q67901. Manipulation of the enhancer  
 CC can be used either to slow or prevent tumor growth or to promote  
 CC differentiation of specific cell types, e.g. epidermal cells to  
 CC promote hair formation. The complete mouse syndecan gene and its  
 CC encoded protein are given in Q67902 and R55276.  
 SQ Sequence 311 AA;

Query Match 67.2%; Score 41; DB 10; Length 311;  
 Best Local Similarity 55.6%; Pred.No. 2.71e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 96 ekpeesepv 104  
 QY 2 ENPDSSEPV 10

RESULT 12  
 ID R87001 standard; Protein; 311 AA.  
 AC R87001;  
 DT 13-AUG-1996 (first entry)  
 DE Mouse syndecan-1.  
 KW Syndecan-1; ectodomain; tumour; cancer; therapy; proteoglycan;  
 KW cell differentiation.  
 OS Mus sp.  
 PN WO9534316-A1.  
 PD 21-DEC-1995.  
 PF 13-JUN-1995; F10344.  
 PR 13-JUN-1994; US-258862.  
 PA (JALK/) JALKANEN M.  
 PA (MALI/) MALI M.  
 PI Jalkanen M, Mali M;  
 DR WPI: 96-049416/05.  
 DR N-PSDB: T08126.  
 DT New syndecan ectodomain compsn. to reduce tumour growth - for  
 PT delivery to extracellular environment for suppressing tumour growth  
 PT in malignant and non-malignant tumours  
 PS Disclosure; Fig 2; 49pp; English.  
 CC Mouse syndecan-1 (R87001) is a cell surface proteoglycan. Its  
 CC ectodomain is capable of normalising the growth rate and  
 CC differentiation state of malignant cells. By providing the ectodomain  
 CC to a cell's extracellular environment, tumours characterised by a  
 CC loss of syndecan-1 can be treated. Expression of the ectodomain in  
 CC transfected S115 cells normalised malignant growth of the cells. The  
 CC ectodomain was also demonstrated to suppress tumour in vivo growth.  
 SQ Sequence 311 AA;

Query Match 67.2%; Score 41; DB 17; Length 311;  
 Best Local Similarity 55.6%; Pred.No. 2.71e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 96 ekpeesepv 104  
 QY 2 ENPDSSEPV 10

RESULT 13  
 ID W18580 standard; Protein; 919 AA.  
 AC W18580;  
 DT 19-SEP-1997 (first entry)  
 DE Potato alpha-glucosidase.  
 KW Alpha-glucosidase; potato; starch; flavour; transgenic plant.  
 OS Solanum tuberosum cv. Record.  
 PN WO9724448-A1.  
 PD 10-JUL-1997.  
 PF 24-DEC-1996; G03239.  
 PR 28-DEC-1995; GB-026613.  
 PA (NICK-) NICKERSON BIOCHEM LTD.  
 PI Davies HV, Taylor MA;  
 DR WPI: 97-363683/33.  
 DR N-PSDB; T69925.  
 DT Sense and antisense plant alpha-glucosidase nucleic acids - useful  
 PT for modifying alpha-glucosidase activity, e.g. for enhancing plant  
 PT tissue flavour, producing novel starch, to control pre-sprouting,  
 PT etc.  
 PS Example 1; Fig 2; 40pp; English.  
 CC A 105.4 kDa alpha-glucosidase (W18580) is encoded by an isolated  
 CC nucleic acid (T69925) of potato cv. Record. This isolated nucleic  
 CC acid can be used to modify alpha-glucosidase activity in plants or  
 CC transformed microorganisms. Activity is enhanced through the use  
 CC of sense nucleic acids, thereby providing more efficient breakdown  
 CC of starch (useful in malting and brewing), novel starches, or  
 CC enhanced flavour to plant tissues such as fruit. Inhibition of  
 CC alpha-glucosidase activity through the use of antisense sequences  
 CC can inhibit starch breakdown and to control pre-sprouting and dry  
 CC weight/starch loss in cooking, storage and processing.  
 SQ Sequence 919 AA;

Query Match 67.2%; Score 41; DB 23; Length 919;  
 Best Local Similarity 55.6%; Pred.No. 2.71e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 70 eenpeseq 78  
 QY 1 QENPDSSEP 9

RESULT 14  
 ID W42736 standard; Peptide; 10 AA.  
 AC W42736;  
 DT 27-APR-1998 (first entry)  
 DE Antigenic decapeptide derived from the human Sp17 protein.  
 KW Sp17; sperm autoantigen; antigen; antigenic epitope; fertility;  
 KW immunocastration; immunocontraceptive vaccine; autoimmune fertility.  
 OS Synthetic.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Modified\_site 1 /note= "N-terminal acetyl"  
 FT WO9739020-A2.  
 PN 23-OCT-1997.  
 PD 11-APR-1997; U06489.  
 PR 15-APR-1996; US-632535.  
 PA (UYNC-) UNIV NORTH CAROLINA.  
 PI Lea I, Orand MG, Widgren EE;  
 DR WPI: 97-526394/48.  
 PT Sperm protein Sp17 antigenic peptide(s) - used as  
 PT immuno-contraceptive to reduce fertility  
 PS Example 14; Page 96; 160pp; English.  
 CC Peptides W42726-58 are a series of N-terminal acetylated sequential

CC decapeptides corresponding to fragments of the human Sp17 protein.  
 CC Mimotope analysis was performed using the peptide pinblock method.  
 CC Peptides comprising a contiguous segment of the amino acids of Sp17  
 CC have been found to be antigenic. They induce antibodies which recognise  
 CC sperm, and inhibit fertilisation. The peptides can be used in  
 CC immunocontraceptive methods as a immunocontraceptive vaccine to reduce  
 CC the fertility of animals, particularly in females. The presence of the  
 CC antigenic peptides can be detected to diagnose autoimmune fertility in  
 CC both male and female subjects.  
 SQ Sequence 10 AA;

Query Match 65.6%; Score 40; DB 27; Length 10;  
 Best Local Similarity 55.6%; Pred. No. 3.43e+02;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 2 qepekssdp 10  
 || | : | : |  
 QY 1 QENPDSSEP 9

RESULT 15  
 ID W42824 standard; Peptide; 10 AA.  
 AC W42824;  
 DT 27-APR-1998 (first entry)  
 DE Antigenic decapeptide derived from the human Sp17 protein.  
 KW Sp17; sperm autoantigen; antigen; antigenic epitope; fertility;  
 KW immunocontraception; immunocontraceptive vaccine; autoimmune fertility.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN W09739020-A2.  
 PD 23-OCT-1997.  
 PF 11-APR-1997; U06489.  
 PR 15-APR-1996; US-632535.  
 PA (UINC-) UNIV NORTH CAROLINA.  
 PI Lea I, ORAND MG, WIDGREN EE;  
 DR WPI; 97-526394/48.  
 PT Sperm protein Sp17 antigenic peptide(s) - used as  
 PT immunocontraceptive to reduce fertility  
 PS Example 25; Table 5; 160pp; English.  
 CC Peptides W42774-867 are a series of sequential decapeptides corresponding  
 CC to fragments of the human Sp17 protein. Each decapeptide is shifted along  
 CC the Sp17 sequence by one amino acid. W42800-820 are derived from the  
 CC human Sp17 sequence between amino acids 40-80 inclusive, while W42849-67  
 CC are between amino acids 114-149 inclusive. The peptides contain antigenic  
 CC epitopes in mice, rabbits, non-human primates and humans. They induce  
 CC antibodies which recognise sperm, and inhibit fertilisation. The peptides  
 CC can be used in immunocontraceptive methods as a immunocontraceptive  
 CC vaccine to reduce the fertility of animals, particularly in females. The  
 CC presence of the antigenic peptides can be detected to diagnose autoimmune  
 CC fertility in both male and female subjects.  
 SQ Sequence 10 AA;

Query Match 65.6%; Score 40; DB 28; Length 10;  
 Best Local Similarity 55.6%; Pred. No. 3.43e+02;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1 qepekssdp 9  
 || | : | : |  
 QY 1 QENPDSSEP 9

Search completed: Thu May 13 15:44:32 1999  
 Job time : 18 secs.

\*\*\*\*\*  
WIRE  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu May 13 15:41:32 1999; MasPar time 3.79 Seconds  
Tabular output not generated. 78.744 Million cell updates/sec

Title: >US-09-040-485-8  
Description: (3-8) from US09040485.pep (2 of 2)  
Perfect Score: 41  
Sequence: 1 PTGEPQ 6

Scoring table: PAM 150  
Gap 15

Searched: 165420 seqs, 49795644 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl6  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle  
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 17.595; Variance 16.811; scale 1.047

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description             | Pred. No. |
|------------|-------|-------------|--------|----|-------------------------|-----------|
| 1          | 41    | 100.0       | 302    | 5  | O44687                  | 1.45e+00  |
| 2          | 38    | 92.7        | 357    | 13 | T22D1.2 PROTEIN.        | 9.35e+00  |
| 3          | 38    | 92.7        | 357    | 13 | NEUROD-LIKE PROTEIN.    | 9.35e+00  |
| 4          | 38    | 92.7        | 539    | 2  | NEUROD.                 | 9.35e+00  |
| 5          | 38    | 92.7        | 1404   | 4  | PROTEINASE.             | 9.35e+00  |
| 6          | 37    | 90.2        | 102    | 1  | KIAA0375.               | 1.70e+01  |
| 7          | 37    | 90.2        | 107    | 1  | PUTATIVE TRYPTOPHANYL   | 1.70e+01  |
| 8          | 37    | 90.2        | 329    | 14 | STRAIN 1381V VP6 (S3)   | 1.70e+01  |
| 9          | 37    | 90.2        | 329    | 14 | STRAIN 11C812 VP6 (S3)  | 1.70e+01  |
| 10         | 37    | 90.2        | 329    | 14 | STRAIN BTW 17 VP6 (S3)  | 1.70e+01  |
| 11         | 37    | 90.2        | 329    | 14 | STRAIN BTW 13 VP6 (S3)  | 1.70e+01  |
| 12         | 37    | 90.2        | 329    | 14 | VACCINE STRAIN VP6 (S3) | 1.70e+01  |
| 13         | 37    | 90.2        | 329    | 14 | STRAIN 10B80V VP6 (S3)  | 1.70e+01  |
| 14         | 37    | 90.2        | 329    | 14 | STRAIN 17C81W VP6 (S3)  | 1.70e+01  |
| 15         | 37    | 90.2        | 329    | 14 | STRAIN BTW 11 VP6 (S3)  | 1.70e+01  |
| 16         | 37    | 90.2        | 329    | 14 | STRAIN 11UC8 VP6 (S3)   | 1.70e+01  |
| 17         | 37    | 90.2        | 329    | 14 | STRAIN 11UC2 VP6 (S3)   | 1.70e+01  |
| 18         | 37    | 90.2        | 329    | 14 | STRAIN 13B89V VP6 (S3)  | 1.70e+01  |
| 19         | 37    | 90.2        | 329    | 14 | STRAIN 110812 VP6 (S3)  | 1.70e+01  |
| 20         | 37    | 90.2        | 329    | 14 | STRAIN 10080V VP6 (S3)  | 1.70e+01  |
|            |       |             |        |    | STRAIN 10B90Z VP6 (S3)  | 1.70e+01  |

|    |    |      |       |    |        |          |
|----|----|------|-------|----|--------|----------|
| 21 | 37 | 90.2 | 329   | 14 | Q98829 | 1.70e+01 |
| 22 | 37 | 90.2 | 329   | 14 | Q98816 | 1.70e+01 |
| 23 | 37 | 90.2 | 329   | 14 | Q96650 | 1.70e+01 |
| 24 | 37 | 90.2 | 329   | 14 | Q96650 | 1.70e+01 |
| 25 | 37 | 90.2 | 369   | 2  | P73589 | 1.70e+01 |
| 26 | 37 | 90.2 | 494   | 2  | Q55721 | 1.70e+01 |
| 27 | 37 | 90.2 | 505   | 2  | P94662 | 1.70e+01 |
| 28 | 37 | 90.2 | 572   | 10 | Q64549 | 1.70e+01 |
| 29 | 36 | 87.8 | 772   | 3  | Q12661 | 1.70e+01 |
| 30 | 36 | 87.8 | 773   | 3  | Q01738 | 3.07e+01 |
| 31 | 36 | 87.8 | 1584  | 5  | Q93791 | 3.07e+01 |
| 32 | 35 | 85.4 | 93    | 4  | Q15215 | 5.47e+01 |
| 33 | 35 | 85.4 | 234   | 11 | Q35859 | 5.47e+01 |
| 34 | 35 | 85.4 | 297   | 4  | Q16038 | 5.47e+01 |
| 35 | 35 | 85.4 | 303   | 2  | Q53566 | 5.47e+01 |
| 36 | 35 | 85.4 | 358   | 4  | Q08805 | 5.47e+01 |
| 37 | 35 | 85.4 | 415   | 11 | Q62848 | 5.47e+01 |
| 38 | 35 | 85.4 | 441   | 2  | P72934 | 5.47e+01 |
| 39 | 35 | 85.4 | 472   | 2  | P95423 | 5.47e+01 |
| 40 | 35 | 85.4 | 742   | 5  | Q61708 | 5.47e+01 |
| 41 | 35 | 85.4 | 1735  | 4  | Q43152 | 5.47e+01 |
| 42 | 35 | 85.4 | 2135  | 5  | Q61077 | 5.47e+01 |
| 43 | 35 | 85.4 | 2611  | 2  | Q68487 | 5.47e+01 |
| 44 | 35 | 85.4 | 6420  | 2  | P95814 | 5.47e+01 |
| 45 | 35 | 85.4 | 13055 | 5  | Q09165 | 5.47e+01 |

ALIGNMENTS

RESULT 1  
ID O44687 PRELIMINARY; PRT; 302 AA.  
AC O44687;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE T22D1.2 PROTEIN.  
GN T22D1.2  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE; 94150718;  
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,  
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWKNEEN R., SWALDON N., SMITH A.,  
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,  
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.;  
RA NATURE 369:32-38(1994)  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX GEISEL C., BRADSHAW H., HAWKINS M.;  
RA SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX WATERSTON R.;  
RA SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.  
DR EMBL; AF039052; G2736513; -  
SO SEQUENCE 302 AA; 30355 MW; 6344F215 CRC32;

Query Match 100.0%; Score 41; DB 5; Length 302;  
Best Local Similarity 100.0%; Pred. No. 1.45e+00;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 47 PTGEPQ 52  
|||||

QY 3 PTGEPQ 8

RESULT 2  
ID P79765 PRELIMINARY; PRT; 357 AA.

AC P79765;  
DT 01-MAY-1997 (TREMREL. 03, CREATED)  
DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
DE NEUROD-LIKE PROTEIN.

OS GALLUS GALLUS (CHICKEN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;  
OC GALLIFORMES.

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WHITE LEHORN; TISSUE-RETINA;  
RA ROTOCIL T., MATTER-SADZINSKI L., ALLIOD C., BALLIVET M., MATTER J.M.;  
RL DEVELOPMENT 124:3263-3272(1997).

DR EMBL; Y09596; E283389; -;  
DR PFAM; PF00010; HLH.  
SQ SEQUENCE 357 AA; 38809 MW; 04A96823 CRC32;

Query Match 92.7%; Score 38; DB 13; Length 357;  
Best Local Similarity 83.3%; Pred. No. 9.35e+00;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 10 PAGEPQ 15

I:||||

QY 3 PTGEPQ 8

RESULT 3  
ID O73826 PRELIMINARY; PRT; 357 AA.

AC O73826;  
DT 01-AUG-1998 (TREMREL. 07, CREATED)  
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
DE NEUROD.

OS GALLUS GALLUS (CHICKEN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;  
OC GALLIFORMES.

RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-WHITE LEHORN;  
RA YAN R.-T., WANG S.-Z.;  
RL J. NEUROBIOL. 0:0-0(1998).

DR EMBL; AF060885; G3094020; -;  
SQ SEQUENCE 357 AA; 38858 MW; 7D406FF3 CRC32;

Query Match 92.7%; Score 38; DB 13; Length 357;  
Best Local Similarity 83.3%; Pred. No. 9.35e+00;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 10 PAGEPQ 15

I:||||

QY 3 PTGEPQ 8

RESULT 4  
ID Q54398 PRELIMINARY; PRT; 539 AA.

AC Q54398;  
DT 01-NOV-1996 (TREMREL. 01, CREATED)  
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)  
DE PROTEINASE.

GN SLPD.  
OS STREPTOMYCES LIVIDANS.  
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=66;  
RX MEDLINE; 96042075.

RA BINNIE C., BUTLER M.J., APHALE J.S., BOURGAULT R., DIZONNO M.A.,  
RA KRYGSMAN P., LIAO L., WALCZYK E., MALEK L.T.;  
RL J. BACTERIOL. 177:6033-6040(1995).

DR EMBL; L42758; G940303; -;  
DR PFAM; PF00561; abhydrolase.

SQ SEQUENCE 539 AA; 57033 MW; 6EAA9AC1 CRC32;

Query Match 92.7%; Score 38; DB 2; Length 539;  
Best Local Similarity 83.3%; Pred. No. 9.35e+00;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 457 PTGEPH 462

I:||||

QY 3 PTGEPQ 8

RESULT 5  
ID O15080 PRELIMINARY; PRT; 1404 AA.

AC O15080;  
DT 01-JAN-1998 (TREMREL. 05, CREATED)  
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)  
DE KIAA0375.

GN KIAA0375.  
OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.

RN [1]  
RP SEQUENCE FROM N.A.

RC TISSUE-BRAIN;  
RX MEDLINE; 97349984.

RA NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N.,  
RA TANAKA A., KOTANI H., NOMURA N., OHARA O.;  
RL DNA RES. 4:141-150(1997).

DR EMBL; AB002373; D1021671; -;  
SQ SEQUENCE 1404 AA; 149264 MW; 675DE431 CRC32;

Query Match 92.7%; Score 38; DB 4; Length 1404;  
Best Local Similarity 83.3%; Pred. No. 9.35e+00;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 746 PSGEPO 751

I:||||

QY 3 PTGEPQ 8

RESULT 6  
ID O07119 PRELIMINARY; PRT; 102 AA.

AC O07119;  
DT 01-JUL-1997 (TREMREL. 04, CREATED)  
DT 01-JUL-1997 (TREMREL. 04, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
DE PUTATIVE TRYPTOPHYL AMINOACYL TRNA SYNTHETASE (FRAGMENT).  
GN TRPS.

OS HALOBACTERIUM VOLCANII (HALOFERAX VOLCANII).

OC ARCHAEABACTERIA; EURYARCHAEOTA; HALOBACTERIALES; HALOBACTERIACEAE.  
RN [1]

RP SEQUENCE FROM N.A.

RA KLEMAN-LEYER K., ARMERUSTER D.W., DANIELS C.J.;

RL CELL 89:839-847(1997).

DR EMBL; AF001578; G2160793; -;

KW AMINOACYL-TRNA SYNTHASE.

FT NON-TER 102 102

SQ SEQUENCE 102 AA; 11442 MW; E13C318F CRC32;

Query Match 90.2%; Score 37; DB 1; Length 102;  
Best Local Similarity 66.7%; Pred. No. 1.70e+01;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 94 PTGDPH 99

I:||||

QY 3 PTGEPQ 8

```
RESULT 7
ID Q98824 PRELIMINARY; PRT; 329 AA.
AC Q98824;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE STRAIN 13881V VP6 (S3).
GN S3.
OS BLUETONGUE VIRUS.
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=13881V;
RA DE MATTOS C.C.; DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D.,
RA YILMA T., OSBURN B.I.;
RL J. VIROL. 70:5735-5739(1996).
DR EMBL; U55788; G1497953; -.
DR EMBL; U55789; G1497955; -.
DR EMBL; U55794; G1497965; -.
SQ SEQUENCE 329 AA; 35556 MW; 7A06D945 CRC32;

Query Match 90.2%; Score 37; DB 14; Length 329;
Best Local Similarity 66.7%; Pred. No. 1.70e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
QY 3 PTGEQ 8

RESULT 8
ID Q96652 PRELIMINARY; PRT; 329 AA.
AC Q96652;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE STRAIN 11C81Z VP6 (S3).
GN S3.
OS BLUETONGUE VIRUS.
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=11C81Z;
RA DE MATTOS C.C.; DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D.,
RA YILMA T., OSBURN B.I.;
RL J. VIROL. 70:5735-5739(1996).
DR EMBL; U55785; G1497947; -.
SQ SEQUENCE 329 AA; 35273 MW; CE93B1B7 CRC32;

Query Match 90.2%; Score 37; DB 14; Length 329;
Best Local Similarity 66.7%; Pred. No. 1.70e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
QY 3 PTGEQ 8

RESULT 9
ID Q96659 PRELIMINARY; PRT; 329 AA.
AC Q96659;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE STRAIN BT17 VP6 (S3).
GN S3.
OS BLUETONGUE VIRUS.
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN-BTV 17;
RX MEDLINE; 96357114.
RA DE MATTOS C.C.; DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D.,
RA YILMA T., OSBURN B.I.;
RL J. VIROL. 70:5735-5739(1996).
DR EMBL; U55798; G1497973; -.
SQ SEQUENCE 329 AA; 35659 MW; 97FE1B22 CRC32;

Query Match 90.2%; Score 37; DB 14; Length 329;
Best Local Similarity 66.7%; Pred. No. 1.70e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
QY 3 PTGEQ 8

RESULT 10
ID Q96658 PRELIMINARY; PRT; 329 AA.
AC Q96658;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE STRAIN BT13 VP6 (S3).
GN S3.
OS BLUETONGUE VIRUS.
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BTV 13;
RA DE MATTOS C.C.; DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D.,
RA YILMA T., OSBURN B.I.;
RL J. VIROL. 70:5735-5739(1996).
DR EMBL; U55797; G1497971; -.
SQ SEQUENCE 329 AA; 35542 MW; B581CF73 CRC32;

Query Match 90.2%; Score 37; DB 14; Length 329;
Best Local Similarity 66.7%; Pred. No. 1.70e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
QY 3 PTGEQ 8

RESULT 11
ID Q96661 PRELIMINARY; PRT; 329 AA.
AC Q96661;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE VACCINE STRAIN VP6 (S3).
GN S3.
OS BLUETONGUE VIRUS.
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VACCINE STRAIN;
RA DE MATTOS C.C.; DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D.,
RA YILMA T., OSBURN B.I.;
RL J. VIROL. 70:5735-5739(1996).
DR EMBL; U55800; G1497977; -.
SQ SEQUENCE 329 AA; 35517 MW; 651408E6 CRC32;

Query Match 90.2%; Score 37; DB 14; Length 329;
Best Local Similarity 66.7%; Pred. No. 1.70e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
QY 3 PTGEQ 8
```

```
RESULT 12
ID Q96648 PRELIMINARY; PRT; 329 AA.
AC Q96648;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE STRAIN 1080Y VP6 (S3).
GN S3.
OS BLUETONGUE VIRUS.
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-1080Y;
RX MEDLINE; 96357114.
RA DE MATTOS C.C., DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D.,
RA YILMA T., OSBURN B.I.;
RL J. VIROL. 70:5735-5739(1996).
DR EMBL; U55778; G1497933; -.
SQ SEQUENCE 329 AA; 35585 MW; 95C25B6C CRC32;

Query Match 90.2%; Score 37; DB 14; Length 329;
Best Local Similarity 66.7%; Pred. No. 1.70e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
|||:|
QY 3 PTGEPQ 8

RESULT 13
ID Q96656 PRELIMINARY; PRT; 329 AA.
AC Q96656;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE STRAIN 17C81W VP6 (S3).
GN S3.
OS BLUETONGUE VIRUS.
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-17C81W;
RX MEDLINE; 96357114.
RA DE MATTOS C.C., DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D.,
RA YILMA T., OSBURN B.I.;
RL J. VIROL. 70:5735-5739(1996).
DR EMBL; U55793; G1497963; -.
SQ SEQUENCE 329 AA; 35535 MW; 5C96CCD9 CRC32;

Query Match 90.2%; Score 37; DB 14; Length 329;
Best Local Similarity 66.7%; Pred. No. 1.70e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
|||:|
QY 3 PTGEPQ 8

RESULT 14
ID Q96657 PRELIMINARY; PRT; 329 AA.
AC Q96657;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE STRAIN BTV 11 VP6 (S3).
GN S3.
OS BLUETONGUE VIRUS.
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BTV 11;
```

```
RX MEDLINE; 96357114.
RA DE MATTOS C.C., DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D.,
RA YILMA T., OSBURN B.I.;
RL J. VIROL. 70:5735-5739(1996).
DR EMBL; U55796; G1497969; -.
SQ SEQUENCE 329 AA; 35629 MW; A645E255 CRC32;

Query Match 90.2%; Score 37; DB 14; Length 329;
Best Local Similarity 66.7%; Pred. No. 1.70e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
|||:|
QY 3 PTGEPQ 8

RESULT 15
ID Q96654 PRELIMINARY; PRT; 329 AA.
AC Q96654;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE STRAIN 11UC8 VP6 (S3).
GN S3.
OS BLUETONGUE VIRUS.
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-11UC8;
RX MEDLINE; 96357114.
RA DE MATTOS C.C., DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D.,
RA YILMA T., OSBURN B.I.;
RL J. VIROL. 70:5735-5739(1996).
DR EMBL; U55787; G1497951; -.
SQ SEQUENCE 329 AA; 35573 MW; 1B421073 CRC32;

Query Match 90.2%; Score 37; DB 14; Length 329;
Best Local Similarity 66.7%; Pred. No. 1.70e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
|||:|
QY 3 PTGEPQ 8

Search completed: Thu May 13 15:42:06 1999
Job time : 34 secs.
```

\*\*\*\*\*  
W P S R E H (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 13 15:28:47 1999; MasPar time 2.03 Seconds  
Tabular output not generated. 132.450 Million cell updates/sec

Title: >US-09-040-485-6  
Description: (1-10) from US09040485.pep  
Perfect Score: 64  
Sequence: 1 APPEDNPVED 10

Scoring table: PAM 150  
Gap 15

Searched: 74019 seqs, 26840295 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot35  
1:swissprot

Statistics: Mean 21.884; Variance 24.812; scale 0.882

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID           | Description            | Pred. No. |
|------------|-------|---------------|--------|--------------|------------------------|-----------|
| 1          | 64    | 100.0         | 757    | 1 ASPH_HUMAN | ASPARTYL/ASPARAGINYL B | 2.44e-04  |
| 2          | 50    | 78.1          | 704    | 1 KPC1_CAEEL | PROTEIN KINASE C-LIKE  | 7.01e-01  |
| 3          | 46    | 71.9          | 562    | 1 LCB2_KLULA | SERINE PALMITOYLTRANSF | 5.49e+00  |
| 4          | 45    | 70.3          | 365    | 1 TRBP_MOUSE | PROTAMINE-1 RNA BINDIN | 9.01e+00  |
| 5          | 45    | 70.3          | 928    | 1 RB_HUMAN   | RETINOBLASTOMA-ASSOCIA | 9.01e+00  |
| 6          | 45    | 70.3          | 932    | 1 SYL_AKCFU  | LEUCYL-TRNA SYNTHETASE | 9.01e+00  |
| 7          | 45    | 70.3          | 2839   | 1 NFI_HUMAN  | NEUROFIBROMIN (NEUROFI | 9.01e+00  |
| 8          | 45    | 70.3          | 2841   | 1 NFI_MOUSE  | NEUROFIBROMIN (NEUROFI | 9.01e+00  |
| 9          | 44    | 68.8          | 36     | 1 PAHO_LARAR | PANCREATIC HORMONE (PA | 1.47e+01  |
| 10         | 44    | 68.8          | 36     | 1 PAHO_STRCA | PANCREATIC HORMONE (PA | 1.47e+01  |
| 11         | 44    | 68.8          | 80     | 1 PAHO_CHICK | PANCREATIC HORMONE PRE | 1.47e+01  |
| 12         | 44    | 68.8          | 122    | 1 YLC3_YEREN | HYPOTHETICAL 13.7 KD P | 1.47e+01  |
| 13         | 44    | 68.8          | 213    | 1 NEUM_CARAU | NEUROMODULIN (AXONAL M | 1.47e+01  |
| 14         | 44    | 68.8          | 262    | 1 TRT3_CHICK | TROPONIN T, FAST SKELE | 1.47e+01  |
| 15         | 44    | 68.8          | 376    | 1 ACTM_APLCA | ACTIN, MUSCLE          | 1.47e+01  |
| 16         | 44    | 68.8          | 395    | 1 YACE_VIBCH | HYPOTHETICAL 44.0 KD P | 1.47e+01  |
| 17         | 44    | 68.8          | 461    | 1 YAU2_SCHPO | HYPOTHETICAL 52.9 KD P | 1.47e+01  |
| 18         | 44    | 68.8          | 647    | 1 DPOG_CHICK | MHC POLYMERASE GAMMA ( | 1.47e+01  |
| 19         | 44    | 68.8          | 979    | 1 RFX1_HUMAN | MHC CLASS II REGULATOR | 1.47e+01  |
| 20         | 43    | 67.2          | 307    | 1 OOX2_ACEAC | UBIQUINOL OXIDASE POLY | 2.37e+01  |
| 21         | 43    | 67.2          | 491    | 1 NWT_CRYNE  | GLYCYLPEPTIDE N-TETRA  | 2.37e+01  |
| 22         | 43    | 67.2          | 506    | 1 GAG_POLYPR | GAG POLYPROTEIN (CORE  | 2.37e+01  |
| 23         | 43    | 67.2          | 507    | 1 GAG_SIVSP  | GAG POLYPROTEIN (CORE  | 2.37e+01  |

|    |    |      |      |              |                        |          |
|----|----|------|------|--------------|------------------------|----------|
| 24 | 43 | 67.2 | 507  | 1 GAG_SIVS4  | GAG POLYPROTEIN (CORE  | 2.37e+01 |
| 25 | 43 | 67.2 | 705  | 1 ICAL_BOVIN | CALPAIN INHIBITOR (CAL | 2.37e+01 |
| 26 | 43 | 67.2 | 966  | 1 SSNG_YEAST | GLUCOSE REPRESSION MED | 2.37e+01 |
| 27 | 43 | 67.2 | 976  | 1 XPR6_YARLI | DIBASIC PROCESSING END | 2.37e+01 |
| 28 | 43 | 67.2 | 1257 | 1 CAML_HUMAN | NEURAL CELL ADHESION M | 2.37e+01 |
| 29 | 43 | 67.2 | 1259 | 1 CAML_RAT   | NEURAL CELL ADHESION M | 2.37e+01 |
| 30 | 43 | 67.2 | 1260 | 1 CAML_MOUSE | NEURAL CELL ADHESION M | 2.37e+01 |
| 31 | 42 | 65.6 | 273  | 1 TRY6_ANOGA | TRYPsin 6 PRECURSOR (E | 3.79e+01 |
| 32 | 42 | 65.6 | 313  | 1 NOD3_RHIME | MODULATION PROTEIN D I | 3.79e+01 |
| 33 | 42 | 65.6 | 403  | 1 CGB3_CHICK | G2/MITOTIC-SPECIFIC CY | 3.79e+01 |
| 34 | 42 | 65.6 | 423  | 1 YPN3_CAEEL | HYPOTHETICAL 47 KD PRO | 3.79e+01 |
| 35 | 42 | 65.6 | 448  | 1 YNO6_YEAST | HYPOTHETICAL 49.7 KD P | 3.79e+01 |
| 36 | 42 | 65.6 | 506  | 1 GAG_SIVM1  | GAG POLYPROTEIN (CORE  | 3.79e+01 |
| 37 | 42 | 65.6 | 547  | 1 RCC_DROME  | REGULATOR OF CHROMOS   | 3.79e+01 |
| 38 | 42 | 65.6 | 742  | 1 CDCH_HALSA | CDCH PROTEIN.          | 3.79e+01 |
| 39 | 42 | 65.6 | 823  | 1 SCH9_YEAST | CAMP-DEPENDENT PROTEIN | 3.79e+01 |
| 40 | 42 | 65.6 | 865  | 1 TOP1_ECOLI | DNA TOPOISOMERASE I (E | 3.79e+01 |
| 41 | 42 | 65.6 | 909  | 1 CNG4_HUMAN | CYCLIC-NUCLEOTIDE-GATE | 3.79e+01 |
| 42 | 42 | 65.6 | 978  | 1 YSX7_CAEEL | HYPOTHETICAL 113.1 KD  | 3.79e+01 |
| 43 | 42 | 65.6 | 1836 | 1 CINA_HUMAN | SODIUM CHANNEL PROTEIN | 3.79e+01 |
| 44 | 42 | 65.6 | 1840 | 1 CINA_RAT   | SODIUM CHANNEL PROTEIN | 3.79e+01 |
| 45 | 42 | 65.6 | 3414 | 1 POLG_LANVT | GENOME POLYPROTEIN (CO | 3.79e+01 |

ALIGNMENTS

RESULT 1  
ID ASPH\_HUMAN STANDARD: PRT: 757 AA.  
AC Q12797;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-HYDROXYLASE).  
DE DIOXYGENASE).  
GN ASPH.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 95121937.  
RA KORIOTH F, GIEFFERS C, FREY J.;  
RL 130:395-399(1994).  
CC -!- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF PROTEINS.  
CC -!- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) = PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).  
CC -!- COFACTOR: IRON.  
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM.  
CC -!- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES TESTED.  
CC -!- PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 274-757) OR 52 KD (AA 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY SIMILARITY).  
EMBL: U03109; E82591; -.  
MIM: 600582; -.  
KW OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR; ENDOPLASMIC RETICULUM.  
FT DOMAIN 1 54  
FT TRANSMEM 55 75  
FT DOMAIN 76 757  
FT DOMAIN 13 20  
FT DOMAIN 323 332  
FT CARBOHYD 452 452  
FT CARBOHYD 705 705  
SQ SEQUENCE 757 AA; 85498 MW; A66AFC24 CRC32;

Query Match 100.0%; Score 64; DB 1; Length 757;  
Best Local Similarity 100.0%; Pred. No. 2.44e-04;



Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 281 APPENDPVED 290  
QY 1 APPENDPVED 10

RESULT 2  
ID KPC1\_CAEEL STANDARD; PRT; 704 AA.  
AC P34722;  
DT 01-FEB-1994 (REL. 28, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE PROTEIN KINASE C-LIKE 1 (EC 2.7.1.-) (PKC).  
GN TPA-1.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE; 95387388.  
RA SANO T., TABUSE Y., NISHIWAKI K., MIWA J.;  
RL J. MOL. BIOL. 251:477-485(1995).  
RN [2]  
RP SEQUENCE OF 148-704 FROM N.A.  
RX MEDLINE; 89186920.  
RA TABUSE Y., NISHIWAKI K., MIWA J.;  
RL SCIENCE 243:1713-1716(1989).  
RN [3]  
RP REVISIONS TO REF. 2.  
RA MIWA J.;

RL SUBMITTED (MAR-1993) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN  
CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS  
CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.  
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS (TPA-1A AND TPA-1B) ARE PRODUCED  
CC BY ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS  
CC THAT OF TPA-1A.  
CC -!- SIMILARITY: CONTAINS TWO COPIES OF THE ZINC-DEPENDENT PHORBOL-  
CC ESTER AND DAG BINDING DOMAIN.  
CC DR EMBL; D49525; G1217584; -.  
CC DR EMBL; D49525; G1217585; -.  
CC DR EMBL; D14815; G303529; -.  
CC DR HSP; P05132; 1ATP.  
CC DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
CC DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
CC DR PROSITE; PS00479; DAG\_PE\_BINDING\_DOMAIN; 2.  
CC DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
CC KW TRANSFERASE: SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING; ZINC;  
KW PHORBOL-ESTER BINDING; REPEAT; ALTERNATIVE SPLICING; PHOSPHORYLATION.  
FT DOMAIN 166 215 PHORBOL-ESTER AND DAG BINDING.  
FT DOMAIN 238 287 PHORBOL-ESTER AND DAG BINDING.  
FT DOMAIN 375 634 PROTEIN KINASE.  
FT NP\_BIND 381 389 ATP (BY SIMILARITY).  
FT BINDING 404 404 ATP (BY SIMILARITY).  
FT ACT\_SITE 499 499 BY SIMILARITY.  
FT MOD\_RES 89 89 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
FT MOD\_RES 139 139 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
FT MOD\_RES 324 324 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
FT VARSPLIC 1 137 MISSING (IN TPA-1B).  
SQ SEQUENCE 704 AA; 80298 MW; 84720228 CRC32;

Query Match 78.1%; Score 50; DB 1; Length 704;  
Best Local Similarity 70.0%; Pred. No. 7.01e-01;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 127 APDDHPVED 136  
QY 1 APPENDPVED 10

RESULT 3  
ID LCB2\_KLULA STANDARD; PRT; 562 AA.

AC P48241;  
DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE SERINE PALMITOYLTRANSFERASE 2 (EC 2.3.1.50) (LONG CHAIN BASE  
DE BIOSYNTHESIS PROTEIN 2) (SPT 2).  
GN LCB2.  
OS KLUYVEROMYCES LACTIS (YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAG;  
RX MEDLINE; 97080528.  
RA NAGIEC M.M., LESTER R.L., DICKSON R.C.;  
RL GENE 177:237-241(1996).  
CC -!- CATALYTIC ACTIVITY: PALMITOYL-COA + L-SERINE = COA + 3-DEHYDRO-  
CC D-SPHINGANINE + CO(2).  
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.  
CC -!- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF THE LONG-CHAIN BASE  
CC COMPONENT OF SPHINGOLIPIDS.  
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.  
CC -!- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
CC AMINOTRANSFERASES.  
CC EMBL; U15646; G1001949; -.  
CC DR PROSITE; PS00599; AA\_TRANSFER\_CLASS\_2; 1.  
CC KW TRANSFERASE; ACYLTRANSFERASE; TRANSMEMBRANE; PYRIDOXAL PHOSPHATE.  
CC TRANSMEM 61 81 POTENTIAL.  
CC FT TRANSMEM 356 376 POTENTIAL.  
CC FT BINDING 365 365 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
SQ SEQUENCE 562 AA; 63004 MW; 901ACF38 CRC32;

Query Match 71.9%; Score 46; DB 1; Length 562;  
Best Local Similarity 77.8%; Pred. No. 5.49e+00;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 13 PREDIPLED 21  
QY 2 PPDNPVED 10

RESULT 4  
ID TRBP\_MOUSE STANDARD; PRT; 365 AA.  
AC P97473;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE PROTAMINE-1 RNA BINDING PROTEIN (PRM-1 RNA BINDING PROTEIN).  
GN TARBP2 OR PRBP.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-TESTIS.  
RX MEDLINE; 96220490.  
RA LEE K., FAJARDO M.A., BRAUN R.E.;  
RL MOL. CELL. BIOL. 16:3023-3034(1996).  
CC -!- FUNCTION: BINDS IN VITRO TO THE PRM1 3'UTR. SEEMS TO ACT AS A  
CC GENERAL REPRESSOR OF TRANSLATION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.  
CC -!- SIMILARITY: CONTAINS 3 DBM (DOUBLE-STRANDED RNA-BINDING) DOMAINS.  
CC EMBL; U79962; G1737216; -.  
CC DR MGD; MGI:108201; PRBP.  
KW RNA-BINDING; REPEAT; NUCLEAR PROTEIN.  
FT DOMAIN 79 95 DBM 1.  
FT DOMAIN 208 224 DBM 2.  
FT DOMAIN 342 358 DBM 3.  
SQ SEQUENCE 365 AA; 38789 MW; 0AF5653D CRC32;

Query Match 70.3%; Score 45; DB 1; Length 365;  
Best Local Similarity 75.0%; Pred. No. 9.01e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 117 SPEDTPV 124  
:|||||  
Qy 1 APPENDPV 8

RESULT 5  
ID RE\_HUMAN STANDARD; PRT; 928 AA.  
AC P06400; P78499;  
DT 01-JAN-1988 (REL. 06, CREATED)  
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE RETINOBLASTOMA-ASSOCIATED PROTEIN (P110) (P105-RB) (RB).  
GN RBL.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 88014238.  
RA LEE W.-H., SHEW J.-Y., HONG F.D., SERY T.W., DONOSO L.A., YOUNG L.-J.,  
RA BOOKSTEIN R., LEE E.Y.-H.P.;  
RL NATURE 329:642-645(1987).  
RN [2]  
RP REVISIONS.  
RX MEDLINE; 87149066.  
RA LEE W.-H., BOOKSTEIN R., HONG F.D., YOUNG L.-J., SHEW J.-Y.,  
RA LEE E.Y.-H.P.;  
RL SCIENCE 235:1394-1399(1987).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 88097427.  
RA FRIEND S.H., HOROWITZ J.M., GERBER M.R., WANG X.-F., BOGENMANN E.,  
RA LI F.P., WEINBERG R.A.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 84:9059-9063(1987).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90006771.  
RA MCGEE T.L., YANDELL D.W., DRYJA T.P.;  
RL GENE 80:119-128(1989).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94063891.  
RA TOGUCHIDA J., MCGEE T.L., CICCARELLI J.A., EAGLE J.R., YANDELL D.W.,  
RA DRYJA T.P.;  
RL GENOMICS 17:535-543(1993).  
RN [6]  
RP SEQUENCE FROM N.A.  
RX TISSUE-CARCINOMA;  
RC MEDLINE; 92319557.  
RA HOGG A., ONADIM Z., BAIRD P.N., COWELL J.K.;  
RL ONCOGENE 7:1445-1451(1992).  
RN [7]  
RP SEQUENCE OF 1-45 FROM N.A.  
RX MEDLINE; 89239464.  
RA T'ANG A., WU K.J., HASHIMOTO T., LIU W.Y., TAKAHASHI R., SHI X.H.,  
RA M'HARA K., ZHANG F.H., CHEN Y.Y., DU C., QIAN J., LIN Y.G.,  
RA MURPREE A.L., QIU W.R., THOMPSON T., BENEDICT W.F., FUNG Y.K.T.;  
RL ONCOGENE 4:401-407(1989).  
RN [8]  
RP PHOSPHORYLATION SITES.  
RX MEDLINE; 92097548.  
RA LEES J.A., BUCHKOVICH K.J., MARSHAK D.R., ANDERSON C.W., HARLOW E.;  
RL EMBO J. 10:4279-4290(1991).  
RN [9]  
RP VARIANT LEU-567.  
RX MEDLINE; 90081757.  
RA YANDELL D.W., CAMPBELL T.A., DAYTON S.H., PETERSEN R., WALTON D.,  
RA LITTLE J.B., MCCONKIE-ROSELL A., BUCKLEY E., DRYJA T.;  
RL NEW ENGL. J. MED. 321:1689-1695(1989).  
RN [10]  
RP VARIANT TRP-661.  
RX MEDLINE; 92335261.  
RA ONADIM Z., HOGG A., BAIRD P.N., COWELL J.K.;

PROC. NATL. ACAD. SCI. U.S.A. 89:6177-6181(1992).  
[11]  
RN VARIANTS ASN-480 DEL AND TRP-661.  
RX MEDLINE; 95012220.  
RA LOHMANN D.R., BRANDT B., HOEPFING W., PASSARGE E., HORSTHEMKE B.;  
RL HUM. GENET. 94:349-354(1994).  
RN [12]  
RP VARIANTS RB ASP-137 AND TRP-661.  
RX MEDLINE; 97456418.  
RA LOHMANN D.R., GERICK M., BRANDT B., OELSCHLAEGER U., LORENZ B.,  
RA PASSARGE E., HORSTHEMKE B.;  
RL AM. J. HUM. GENET. 61:282-294(1997).  
CC -!- FUNCTION: PROBABLY ACTS AS A REGULATOR OF OTHER GENES. FORMS A  
CC COMPLEX WITH ADENOVIRUS E1A AND WITH SV40 LARGE T ANTIGEN. ACTS AS  
CC A TUMOR SUPPRESSOR. MAY BIND AND MODULATE FUNCTIONALLY CERTAIN  
CC CELLULAR PROTEINS WITH WHICH T AND E1A COMPETE FOR POCKET BINDING.  
CC POTENT INHIBITOR OF E2F-MEDIATED TRANS-ACTIVATION, INTERACTS  
CC PREFERENTIALLY WITH TRANSCRIPTION FACTOR E2F1.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- PTM: PHOSPHORYLATED FROM S TO M PHASE OF THE CELL CYCLE AND IS  
CC DEPHOSPHORYLATED IN G1. T, BUT NOT E1A, BINDS ONLY TO THE  
CC UNPHOSPHORYLATED FORM.  
CC -!- DISEASE: DELETION OR ALTERATION OF THIS PROTEIN RESULTS IN THE  
CC CHILDHOOD CANCER RETINOBLASTOMA (RB). RB IS A CONGENITAL  
CC MALIGNANT TUMOR THAT ARISES FROM THE NUCLEAR LAYERS OF THE  
CC RETINA. IT OCCURS IN ABOUT 1:20 000 LIVE BIRTHS AND REPRESENTS  
CC ABOUT 2% OF CHILDHOOD MALIGNANCIES. IT IS BILATERAL IN ABOUT 30%  
CC OF CASES. ALTHOUGH MOST RB APPEAR SPORADICALLY, ABOUT 20% ARE  
CC TRANSMITTED AS AN AUTOSOMAL DOMINANT TRAIT WITH INCOMPLETE  
CC PENETRANCE. THE DIAGNOSIS IS USUALLY MADE BEFORE THE AGE OF 2  
CC YEARS WHEN STRABISMUS OR A GREY TO YELLOW REFLEX FROM PUPIL ("CAT  
CC EYE") IS INVESTIGATED.  
CC -!- SIMILARITY: BELONGS TO THE RETINOBLASTOMA PROTEIN (RB) FAMILY.  
CC EMBL; L41870; G793995; -  
DR EMBL; M27866; G521212; -  
DR EMBL; M27845; G521212; JOINED.  
DR EMBL; M27846; G521212; JOINED.  
DR EMBL; M27847; G521212; JOINED.  
DR EMBL; M27849; G521212; JOINED.  
DR EMBL; M27850; G521212; JOINED.  
DR EMBL; M27851; G521212; JOINED.  
DR EMBL; L35146; G521212; JOINED.  
DR EMBL; M27852; G521212; JOINED.  
DR EMBL; M27853; G521212; JOINED.  
DR EMBL; M27854; G521212; JOINED.  
DR EMBL; M27855; G521212; JOINED.  
DR EMBL; M27856; G521212; JOINED.  
DR EMBL; M27857; G521212; JOINED.  
DR EMBL; M27858; G521212; JOINED.  
DR EMBL; M27859; G521212; JOINED.  
DR EMBL; M27860; G521212; JOINED.  
DR EMBL; L35147; G521212; JOINED.  
DR EMBL; M27862; G521212; JOINED.  
DR EMBL; M27863; G521212; JOINED.  
DR EMBL; M27864; G521212; JOINED.  
DR EMBL; M27865; G521212; JOINED.  
DR EMBL; M15400; G190959; JOINED.  
DR EMBL; M28419; G190963; -  
DR EMBL; M33647; G190946; -  
DR EMBL; X16439; G35895; -  
DR EMBL; L41997; G804760; -  
DR EMBL; L41890; G793951; -  
DR EMBL; L41891; G793953; -  
DR EMBL; L41893; G793959; -  
DR EMBL; L41894; G793961; -  
DR EMBL; L41895; G793963; -  
DR EMBL; L41896; G793965; -  
DR EMBL; L41897; G793967; -  
DR EMBL; L41898; G793971; -  
DR EMBL; L41899; G793976; -  
DR EMBL; L11910; G292421; -  
DR PIR; JS0276; RBH0.  
DR PIR; A39947; A39947.

DR TRANSFAC; T00722; -  
 DR MIM; 180200; -  
 DR MIM; 259200; -  
 DR MIM; 239500; -  
 KW TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN;  
 KW PHOSPHORYLATION; ANTI-ONCOGENE; DISEASE MUTATION.  
 FT DOMAIN 10 18 POLY-ALA.  
 FT DOMAIN 20 29 POLY-PRO.  
 FT DOMAIN 373 771 POCKET (BINDS T AND EIA).  
 FT DOMAIN 373 579 DOMAIN A.  
 FT DOMAIN 580 639 SPACER.  
 FT DOMAIN 640 771 DOMAIN B.  
 FT MOD\_RES 249 249 PHOSPHORYLATION (BY CDC2).  
 FT MOD\_RES 252 252 PHOSPHORYLATION (BY CDC2).  
 FT MOD\_RES 373 373 PHOSPHORYLATION (BY CDC2).  
 FT MOD\_RES 807 807 PHOSPHORYLATION (BY CDC2).  
 FT MOD\_RES 811 811 PHOSPHORYLATION (BY CDC2).  
 FT VARIANT 137 137 E -> D (IN UNILATERAL RB).  
 FT VARIANT 480 480 MISSING (IN RB; MILD FORM).  
 FT VARIANT 567 567 S -> L (IN RB).  
 FT VARIANT 661 661 R -> W (IN RB; MILD FORM).  
 SQ SEQUENCE 928 AA; 106159 MW; 0BD3B72E CRC32;

Query Match 70.3%; Score 45; DB 1; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 9.01e+00;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 PPPEEDPEQD 36  
 QY 1 APPEDNPVED 10

RESULT 6  
 ID SYL\_ARCFU STANDARD; PRT; 932 AA.  
 AC O30250;  
 DT 15-JUL-1998 (REL. 36, CREATED)  
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE LEUCYL-TRNA SYNTHETASE (EC 6.1.1.4) (LEUCINE-TRNA LIGASE) (LEURS).  
 GN LEUS OR AF2421.  
 OS ARCHAEoglobus fulgidus.  
 OC ARCHAEABACTERIA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE; 98049343.  
 RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,  
 RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,  
 RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KIRPIDES N.C.,  
 RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,  
 RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,  
 RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,  
 RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,  
 RA COTTON M.D., SPRIGGS T., ARTIACH P., KATNE B.P., SKES S.M.,  
 RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,  
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,  
 RA VENTER J.C.;  
 RL NATURE 390:364-370(1997).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-LEUCINE + TRNA(LEU) = AMP +  
 CC -!- PYROPHOSPHATE + L-LEUCYL-TRNA(LEU).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.  
 DR EMBL; AE001108; G2650668; -;  
 DR TIGR; AF2421; -;  
 DR PROSITE; PS00178; AA-TRNA\_LIGASE\_1;  
 KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.  
 FT SIMILAR 38 48 "HIGH" REGION.  
 FT SIMILAR 630 634 "KMSKS" REGION.  
 FT BINDING 633 633 ATP (BY SIMILARITY).  
 SQ SEQUENCE 932 AA; 108628 MW; 13803D54 CRC32;

Query Match 70.3%; Score 45; DB 1; Length 932;  
 Best Local Similarity 66.7%; Pred. No. 9.01e+00;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 180 PHQNPVED 188  
 QY 2 PPEDNPVED 10  
 RESULT 7  
 ID NFL\_HUMAN STANDARD; PRT; 2839 AA.  
 AC P21359;  
 DT 01-MAY-1991 (REL. 18, CREATED)  
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE NEUROFIBROMIN (NEUROFIBROMATOSIS-RELATED PROTEIN NF-1).  
 GN NFL.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE OF 1-1370 AND 1392-2839 FROM N.A.  
 RX MEDLINE; 92147138.  
 RA MARCHUK D.A., SAULINO A., TAVAKKOL R., SWAROOP M., WALLACE M.R.,  
 RA ANDERSEN L.B., MITCHELL A.L., GUTMANN D.H., BOGUSKI M., COLLINS F.S.;  
 RL GENOMICS 11:931-940(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 93090270.  
 RA BERNARDS A., HAASE V.H., MURTHY A.E., MENON A., HANNIGAN G.E.,  
 RA GUSELLA J.F.;  
 RL DNA CELL BIOL. 11:727-734(1992).  
 RN [3]  
 RP SEQUENCE OF 335-1370 AND 1392-2839 FROM N.A.  
 RX MEDLINE; 90335969.  
 RA XU G., O'CONNELL P., VISKOCHIL D., CANTHON R., ROBERTSON M.,  
 RA CULVER M., DUNN D., STEVENS J., GESTELAND R., WHITE R., WEISS R.;  
 RL CELL 62:599-608(1990).  
 RN [4]  
 RP SEQUENCE OF 1096-1370 AND 1372-1590 FROM N.A.  
 RX MEDLINE; 91029515.  
 RA MARTIN G.A., VISKOCHIL D., BOLLAG G., MCCABE P.C., CROSIER W.J.,  
 RA HAUSBRUCK H., CONROY L., CLARK R., O'CONNELL P., CANTHON R.M.,  
 RA INNIS M., MCCORMICK F.;  
 RL CELL 63:843-849(1990).  
 RN [5]  
 RP SEQUENCE OF 1606-2709 FROM N.A., AND VARIANT PRO-1953.  
 RX MEDLINE; 90304909.  
 RA CANTHON R.M., WEISS R., XU G., VISKOCHIL D., CULVER M., STEVENS J.,  
 RA ROBERTSON M., DUNN D., GESTELAND R., O'CONNELL P., WHITE R.;  
 RL CELL 62:193-201(1990).  
 RN [6]  
 RP SEQUENCE OF 2230-2839 FROM N.A.  
 RX MEDLINE; 90319792  
 RA WALLACE M.R., MARCHUK D.A., ANDERSEN L.B., LETCHER R., ODEH H.M.,  
 RA SAULINO A.M., FOUNTAIN J.W., BRERETON A., NICHOLSON J., MITCHELL A.L.,  
 RA BROWNSTEIN B.H., COLLINS F.S.;  
 RL SCIENCE 249:181-186(1990).  
 RN [7]  
 RP ERRATUM.  
 RX MEDLINE; 91102559.  
 RA WALLACE M.R., MARCHUK D.A., ANDERSEN L.B., COLLINS F.S.;  
 RL SCIENCE 250:1749-1749(1990).  
 RN [8]  
 RP SEQUENCE OF 1168-1566 FROM N.A.  
 RX MEDLINE; 92019823.  
 RA NISHI T., LEE P.S., OKA K., LEVIN V.A., TANASE S., MORINO Y.,  
 RA SAYA H.;  
 RL ONCOGENE 6:1555-1559(1991).  
 RN [9]  
 RP SEQUENCE OF 1371-1391 FROM N.A.  
 RX MEDLINE; 93109335.  
 RA ANDERSEN L.B., BALLESTER R., MARCHUK D.A., CHANG E., GUTMANN D.H.,  
 RA SAULINO A.M., CAMONIS J., WIGLER M., COLLINS F.S.;  
 RL MOL. CELL. BIOL. 13:487-495(1993).

RN [10] FUNCTION.  
 RX MEDLINE; 91029516.  
 RA BALLESTER R., MARCHUK D., BOGUSKI M.S., SAULINO A., LETCHER R.,  
 RA WIGLER M., COLLINS F.S.;  
 RL CELL 63:851-859(1990).  
 RN [11]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE; 95072625.  
 RA UPADHYAYA M., SHAW D.J., HARPER P.S.;  
 RL HUM. MUTAT. 4:83-101(1994).  
 RN [12]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE; 96422425.  
 RA HONG SHEN M., HARPER P.S., UPADHYAYA M.;  
 RL J. MED. GENET. 33:2-17(1996).  
 RN [13]  
 RP VARIANT GLU-1444.  
 RX MEDLINE; 92233464.  
 RA LI Y., BOLLAG G., CLARK R., STEVENS J., CONROY L., FULTS D., WARD K.,  
 RA FRIEDMAN E., SAMOWITZ W., ROBERTSON M., BRADLEY P., MCCORMICK F.,  
 RA WHITE R., CAWTHON R.;  
 RL CELL 69:275-281(1992).  
 RN [14]  
 RP VARIANTS MET-2164 AND ASN-2192.  
 RX MEDLINE; 93258316.  
 RA UPADHYAYA M., SHEN M., CHERRYSON A., FARNHAM J., MAYNARD J.,  
 RA HUSON S.M., HARPER P.S.;  
 RL HUM. MOL. GENET. 1:735-740(1992).  
 RN [15]  
 RP VARIANT HIS-1721--LEU-1733 DUPLICATION.  
 RX MEDLINE; 93304433.  
 RA TASSABEHJI M., STRACHAN T., SHARLAND M., COLLEY A., DONNAI D.,  
 RA HARRIS R., THAKER N.;  
 RL AM. J. HUM. GENET. 53:90-95(1993).  
 RN [16]  
 RP VARIANT MET-991 DEL.  
 RX MEDLINE; 94108439.  
 RA SHEN M.H., HARPER P.S., UPADHYAYA M.;  
 RL HUM. MOL. GENET. 2:1861-1864(1993).  
 RN [17]  
 RP VARIANT NF1 ASN-2387--PHE-2388 DEL.  
 RX MEDLINE; 94362704.  
 RA ABERNATHY C.R., COLMAN S.D., KOUSSEFF B.G., WALLACE M.R.;  
 RL HUM. MUTAT. 3:347-352(1994).  
 RN [18]  
 RP VARIANT NF1 ALA-2631.  
 RX MEDLINE; 96091873.  
 RA UPADHYAYA M., MAYNARD J., OSBORN M., HUSON S.M., PONDER M.,  
 RA PONDER B.A.J., HARPER P.S.;  
 RL J. MED. GENET. 32:706-710(1995).  
 RN [19]  
 RP VARIANT NF1 ARG-629.  
 RX MEDLINE; 96431167.  
 RA GASPARI P., D'AGRUMA L., DE CILLIS G.P., BALESTRAZZI P.,  
 RA MINGARELLI R., ZELANTE L.;  
 RL HUM. GENET. 97:492-495(1996).  
 RN [20]  
 RP VARIANT LS ARG-1035.  
 RX MEDLINE; 96400960.  
 RA WU R., LEGIUS E., ROBBERECHT W., DUMOULIN M., CASSIMAN J.-J.,  
 RA FRYS J.-P.;  
 RL HUM. MUTAT. 8:51-56(1996).  
 RN [21]  
 RP VARIANTS NF1 ARG-844 AND PRO-898.  
 RX MEDLINE; 97295087.  
 RA MAYNARD J., KRAWCZAK M., UPADHYAYA M.;  
 RL HUM. GENET. 99:674-676(1997).  
 RN [22]  
 RP VARIANT NF1 ARG-1952.  
 RX MEDLINE; 97255969.  
 RA HUDSON J., WU C.L., TASSABEHJI M., SUMMERS E.M., SIMON S., SUPER M.,  
 RA DONNAI D., THAKER N.;

HUM. MUTAT. 9:366-367(1997).  
 [23]  
 RN VARIANT NF1 TRP-1611.  
 RX MEDLINE; 97442280.  
 RA UPADHYAYA M., MAYNARD J., OSBORN M., HARPER P.S.;  
 RL HUM. MUTAT. 10:248-250(1997).  
 CC -!- FUNCTION: STIMULATES THE GTPASE ACTIVITY OF RAS. NF1 SHOWS GREATER  
 AFFINITY FOR RAS GAP, BUT LOWER SPECIFIC ACTIVITY. THUS IT MAY BE  
 A REGULATOR OF RAS ACTIVITY.  
 CC -!- DISEASE: THIS PROTEIN IS ASSOCIATED WITH TYPE 1 NEUROFIBROMATOSIS  
 (NF1) (ALSO CALLED VON RECKLINGHAUSEN SYNDROME), THE MOST FREQUENT  
 INHERITED GENETIC DISEASE (ABOUT 1 IN 3000). IT EXHIBITS FULL  
 PENETRANCE AND HIGH MUTATION RATE WITH 30 TO 50% OF NF1 PATIENTS  
 REPRESENTING A NEW MUTATION. AMONG THE MANY CLINICAL FEATURES OF  
 NF1 ARE PATCHES OF SKIN PIGMENTATION (CAFE-AU-LAIT SPOTS), LISCH  
 NODULES OF THE IRIS PERIPHERAL, PERIPHERAL NERVOUS SYSTEM  
 ASSOCIATED TUMORS AND FIBROMATOUS SKIN TUMORS. THE DISEASE  
 DEMONSTRATES A HIGH DEGREE OF PENETRANCE BY AGE 5 YEARS.  
 CC -!- DISEASE: DEFECTS IN NF1 ARE ASSOCIATED WITH WATSON SYNDROME (WS).  
 WS IS CHARACTERIZED BY THE PRESENCE OF PULMONARY STENOSIS,  
 CAFE-AU-LAIT SPOTS, AND MENTAL RETARDATION. WS IS CONSIDERED AS  
 AN ATYPICAL FORM OF NF1.  
 CC -!- DISEASE: DEFECTS IN NF1 ARE ASSOCIATED WITH LEOPARD, AN AUTOSOMAL  
 DOMINANT SYNDROME. LEOPARD (LS) IS AN ACRONYM FOR THE FEATURES OF  
 THIS SYNDROME: L=LENTIGINES ("FRECKLES"), E=ELECTROCARDIOGRAPHIC  
 ABNORMALITIES, O=OCULAR HYPERTELOMISM, P=PULMONARY STENOSIS,  
 A=ABNORMALITIES OF GENITALIA, R=RETARDATION OF GROWTH, AND  
 D=DEAFNESS (SENSORINEURAL). THE MAIN FEATURES OF THE SYNDROME ARE  
 MULTIPLE LENTIGINES IN COMBINATION WITH A CONGENITAL HEART  
 MALFORMATION (PULMONARY STENOSIS, SUBVALVULAR MUSCULAR AORTIC  
 STENOSIS). A CLINICAL OVERLAP EXISTS BETWEEN LS, NF1 AND WS.  
 CC -!- ALTERNATIVE PRODUCTS: TWO FORMS OF NEUROFIBROMIN ARE PRODUCED BY  
 ALTERNATIVE SPLICING OF THE NF1 GENE; TYPE I AND TYPE II (SHOWN  
 HERE) ONLY DIFFERS IN THE INSERTION OF A 21 RESIDUES SEGMENT.  
 CC -!- SIMILARITY: TO OTHER RAS GTPASE-ACTIVATING PROTEINS.  
 DR EMBL; M82814; G189165; -  
 DR EMBL; M89914; G292354; -  
 DR EMBL; M38116; G494225; -  
 DR EMBL; M38108; G494225; JOINED.  
 DR EMBL; M38109; G494225; JOINED.  
 DR EMBL; M38110; G494225; JOINED.  
 DR EMBL; M38111; G494225; JOINED.  
 DR EMBL; M38112; G494225; JOINED.  
 DR EMBL; M38113; G494225; JOINED.  
 DR EMBL; M38114; G494225; JOINED.  
 DR EMBL; M38106; G189170; -  
 DR EMBL; M61213; G189163; -  
 DR EMBL; M38107; G189172; ALT\_SEQ.  
 DR EMBL; M60496; G189158; -  
 DR EMBL; M60915; G189161; -  
 DR EMBL; S51751; G362288; -  
 DR EMBL; D12625; G219940; ALT\_SEQ.  
 DR PIR; A35222; A35222.  
 DR PIR; A35605; A35605.  
 DR PIR; A35879; A35879.  
 DR PIR; JC1277; JC1277.  
 DR MIN; 162200; -  
 DR MIN; 193520; -  
 DR MIN; 151100; -  
 DR PROSITE; PS00509; RAS\_GTPASE\_ACTIV\_1; 1.  
 DR PROSITE; PSS0018; RAS\_GTPASE\_ACTIV\_2; 1.  
 KW GTPASE ACTIVATION; ALTERNATIVE SPLICING; ANTI-ONCOGENE;  
 ...  
 Note: remainder of annotations omitted.  
 Query Match 70.3%; Score 45; DB 1; Length 2839;  
 Best Local Similarity 60.0%; Pred. No. 9.01e+00;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Db 1547 GPPCHKPVAD 1556  
 Qy 1 APPENDPVED 10

```

RESULT 8
ID NFL_MOUSE STANDARD; PRT; 2841 AA.
AC Q04690; Q61956; Q61957;
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE NEUROFIBROMIN (NEUROFIBROMATOSIS-RELATED PROTEIN NF-1).
GN NFL
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-BRAIN;
RX MEDLINE; 93357730.
RA BERNARDS A., SNIDJERS A.J., HANNIGAN G.E., MURTHY A.E., GUSELLA J.F.;
RL HUM. MOL. GENET. 2:645-650(1993).
RN [2]
RP SEQUENCE OF 1178-1555 FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE; 95047432.
RA MANTANI A., MAKASUGI S., YOKOTA Y., ABE K., USHIO Y., YAMAMURA K.;
RL GENE 148:245-251(1994).
RN [3]
RP SEQUENCE OF 1950-2568 FROM N.A.
RX MEDLINE; 90384569.
RA BUCHBERG A.M., CLEVELAND L.S., JENKINS N.A., COPELAND N.G.;
RL NATURE 347:291-294(1990).
CC -!- FUNCTION: STIMULATES THE GTPASE ACTIVITY OF RAS. NFL SHOWS GREATER
AFFINITY FOR RAS GAP, BUT LOWER SPECIFIC ACTIVITY. THUS IT MAY BE
A REGULATOR OF RAS ACTIVITY.
CC -!- TISSUE SPECIFICITY: TYPE I IS EXPRESSED PREDOMINANTLY IN BRAIN,
SPINAL CORD AND TESTIS. TYPE II IS EXPRESSED PREDOMINANTLY IN
ADRENAL GLAND, KIDNEY, OVARY AND LUNG. TYPE III IS EXPRESSED
PREDOMINANTLY IN ADRENAL GLAND AND TYPE IV IS EXPRESSED
MAINLY IN THE TESTIS.
CC -!- ALTERNATIVE PRODUCTS: FOUR FORMS OF THE PROTEIN (TYPES I, II,
III AND IV) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME
GENE. THE SEQUENCE SHOW HERE IS THAT OF TYPE II.
CC -!- SIMILARITY: TO OTHER RAS GTPASE-ACTIVATING PROTEINS.
DR EMBL; L10369; G309451; -
DR EMBL; L10367; G309451; JOINED.
DR EMBL; L10368; G309451; JOINED.
DR EMBL; L10370; G309453; -
DR EMBL; X54924; G930191; -
DR EMBL; D30730; G577638; -
DR EMBL; D30731; G577640; -
DR MGI; 97306; NFL.
DR PROSITE; PS00509; RAS_GTPASE_ACTIV_1; 1.
DR PROSITE; PS50018; RAS_GTPASE_ACTIV_2; 1.
KW GTPASE ACTIVATION; ALTERNATIVE SPLICING.
FT DOMAIN 1237 1453
FT VARSPPLIC 1373 1393
FT VARSPPLIC 1394 1406
FT VARSPPLIC 1407 2841
FT VARSPPLIC 2841 AA; 319591 MW; A7AA76F4 CRC32;
SQ SEQUENCE 2841 AA; 70.3%; Score 45; DB 1; Length 2841;
Query Match 70.3%; Score 45; DB 1; Length 2841;
Best Local Similarity 60.0%; Pred. No. 9.01e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1549 GPEHPKPVAD 1558
:|||||
QY 1 APPEDNPVED 10

RESULT 9
ID PAHO_LARAR STANDARD; PRT; 36 AA.
AC P41337;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
OS GALLUS GALLUS (CHICKEN), AND MELEAGRIS GALLOPAVO (COMMON TURKEY).

```

```

DE PANCREATIC HORMONE (PANCREATIC POLYPEPTIDE) (PP).
OS LARUS ARGENTATUS (HERRING GULL).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC CHARADRIIFORMES.
RN [1]
RP SEQUENCE.
RC TISSUE=PANCREAS;
RX MEDLINE; 94229519.
RA BARTON C.L., SHAW C., HALTON D.W., THIM L.;
RL GEN. COMP. ENDOCRINOL. 93:255-259(1994).
CC -!- FUNCTION: PANCREATIC HORMONE IS SYNTHESIZED IN PANCREATIC ISLETS
OF LANGERHANS AND ACTS AS A REGULATOR OF PANCREATIC AND
GASTROINTESTINAL FUNCTIONS.
CC -!- SIMILARITY: BELONGS TO THE PP FAMILY (PANCREATIC HORMONE,
NEUROPEPTIDE Y, PEPTIDE YY).
DR PROSITE; PS00265; PANCREATIC_HORMONE; 1.
KW HORMONE; AMIDATION; PANCREAS.
FT MOD_RES 36 36
SQ SEQUENCE 36 AA; 4237 MW; 2AF9A777 CRC32;

Query Match 68.8%; Score 44; DB 1; Length 36;
Best Local Similarity 66.7%; Pred. No. 1.47e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 8 PGDDAPVED 16
|:|:|
QY 2 PPEDNPVED 10

RESULT 10
ID PAHO_SYRCA STANDARD; PRT; 36 AA.
AC P11967;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE PANCREATIC HORMONE (PANCREATIC POLYPEPTIDE) (PP).
OS STRUTHIO CAMELUS (OSTRICH).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES;
OC PALAEOGNATHAE; STRUTHIONIFORMES.
RN [1]
RP SEQUENCE.
RX MEDLINE; 87307111.
RA LITTHAUER D., OELOFSEN W.;
RL INT. J. PEPT. PROTEIN RES. 29:739-745(1987).
CC -!- FUNCTION: PANCREATIC HORMONE IS SYNTHESIZED IN PANCREATIC ISLETS
OF LANGERHANS AND ACTS AS A REGULATOR OF PANCREATIC AND
GASTROINTESTINAL FUNCTIONS.
CC -!- SIMILARITY: BELONGS TO THE PP FAMILY (PANCREATIC HORMONE,
NEUROPEPTIDE Y, PEPTIDE YY).
DR PIR; A28578; A28578.
DR HSSP; P01306; 1PPT.
DR PROSITE; PS00265; PANCREATIC_HORMONE; 1.
KW HORMONE; AMIDATION; PANCREAS.
FT MOD_RES 36 36
SQ SEQUENCE 36 AA; 4209 MW; 45F8AEDB CRC32;

Query Match 68.8%; Score 44; DB 1; Length 36;
Best Local Similarity 66.7%; Pred. No. 1.47e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 8 PGDDAPVED 16
|:|:|
QY 2 PPEDNPVED 10

RESULT 11
ID PAHO_CHICK STANDARD; PRT; 80 AA.
AC P01306;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PANCREATIC HORMONE PRECURSOR (PANCREATIC POLYPEPTIDE) (PP).
OS GALLUS GALLUS (CHICKEN), AND MELEAGRIS GALLOPAVO (COMMON TURKEY).

```

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;  
 CC GALLIFORMES.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RX SPECIES=CHICKEN; TISSUE=LIVER;  
 RX MEDLINE; 93366173.  
 RA NATA K., SUGIMOTO T., KOHRI K., HIDAKA H., HATTORI E., YAMAMOTO H.,  
 RA YONEKURA H., OKAMOTO H.;  
 RL GENE 130:183-189(1993).  
 RN [2]  
 RP SEQUENCE OF 26-61.  
 RC SPECIES=CHICKEN;  
 RX MEDLINE; 76069270.  
 RA KIMMEL J.R., HAYDEN L.J., POLLOCK H.G.;  
 RL J. BIOL. CHEM. 250:9369-9376(1975).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.4 ANGSTROMS).  
 RC SPECIES=TURKEY;  
 RX MEDLINE; 84179397.  
 RA GLOVER I., MANEUF I., PITTS J., WOODS S., MOSS D., TICKLE I.,  
 RA BLUNDELL T.L.;  
 RL BIOPOLYMERS 22:293-304(1983).  
 CC -!- FUNCTION: PANCREATIC HORMONE IS SYNTHESIZED IN PANCREATIC ISLETS  
 OF LANGERHANS AND ACTS AS A REGULATOR OF PANCREATIC AND  
 CC GASTROINTESTINAL FUNCTIONS.  
 CC -!- SIMILARITY: BELONGS TO THE PP FAMILY (PANCREATIC HORMONE,  
 CC NEUROPEPTIDE Y, PEPTIDE YY).  
 DR EMBL; D13761; G391634; -.  
 DR EMBL; D13760; G391646; -.  
 DR PIR; A01575; PCCH.  
 DR PDB; 1PPT; 15-OCT-91.  
 DR PROSITE; PS00265; PANCREATIC HORMONE; 1.  
 KW HORMONE; CLEAVAGE ON PAIR OF BASIC RESIDUES; PANCREAS; SIGNAL;  
 KW AMIDATION; 3D-STRUCTURE.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 61 PANCREATIC HORMONE.  
 FT MOD\_RES 61 61 AMIDATION (G-62 PROVIDE AMIDE GROUP).  
 FT CONFLICT 47 48 ND -> DN (IN REF. 2).  
 FT TURN 35 36  
 FT HELIX 39 56  
 FT TURN 57 58  
 SQ SEQUENCE 80 AA; 8773 MW; DE642B29 CRC32;  
 Query Match 68.8%; Score 44; DB 1; Length 80;  
 Best Local Similarity 66.7%; Pred. No. 1.47e+01;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Db 33 PGDDAPVED 41  
 QY 2 PPDNPVED 10  
 RESULT 12  
 ID YLC3\_YEREN STANDARD; PRT; 122 AA.  
 AC P21308;  
 DT 01-MAY-1991 (REL. 18, CREATED)  
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 13.7 KD PROTEIN IN LCRC 3'REGION (ORF3).  
 OS YERSINIA ENTEROCOLITICA.  
 OG PLASMID PYV.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 CC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SEROTYPE O:3;  
 RX MEDLINE; 90264308.  
 RA VIITANEN A.-M., TOIVANEN P., SKURNIK M.;  
 RL J. BACTERIOL. 172:3152-3162(1990).  
 CC -!- INDUCTION: TEMPERATURE SEEMS TO PLAY THE MAJOR ROLE IN REGULATION  
 OF TRANSCRIPTION OF THE LCRC-CONTAINING OPERON OF PYV, WHEREAS  
 CC CA(2+) CONCENTRATION HAS ONLY A MODERATE EFFECT AT 37 DEGREES  
 CC CELSIUS, AND NO EFFECT AT ROOM TEMPERATURE.

DR EMBL; M32097; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; D35392; D35392.  
 KW HYPOTHETICAL PROTEIN; PLASMID.  
 SQ SEQUENCE 122 AA; 13741 MW; 0E8B386D CRC32;  
 Query Match 68.8%; Score 44; DB 1; Length 122;  
 Best Local Similarity 62.5%; Pred. No. 1.47e+01;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 34 PPDGHPVE 41  
 QY 2 PPDNPVE 9  
 RESULT 13  
 ID NEUM\_CARAU STANDARD; PRT; 213 AA.  
 AC P17691;  
 DT 01-AUG-1990 (REL. 15, CREATED)  
 DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE NEUROMODULIN (AXONAL MEMBRANE PROTEIN GAP-43) (PP46) (B-50) (PROTEIN  
 DE FL) (CALMODULIN-BINDING PROTEIN P-57).  
 OS CARASSIUS AURATUS (GOLDFISH).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNATHOSTOMATA;  
 OC OSTEICHTHYES; ACTINOPTERYGII; CYPRINIFORMES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 90380372.  
 RA LABATE M.E., SKENE J.H.P.;  
 RL NEURON 3:299-310(1989).  
 CC -!- FUNCTION: THIS PROTEIN IS ASSOCIATED WITH NERVE GROWTH. IT IS A  
 MAJOR COMPONENT OF THE MOTILE "GROWTH CONES" THAT FORM THE TIPS  
 CC OF ELONGATING AXONS.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF GROWTH CONE AND  
 CC SYNAPTIC PLASMA MEMBRANES.  
 CC -!- PTM: PHOSPHORYLATION OF THIS PROTEIN BY A PROTEIN KINASE C IS  
 SPECIFICALLY CORRELATED WITH CERTAIN FORMS OF SYNAPTIC PLASTICITY.  
 CC -!- BINDS CALMODULIN WITH A GREATER AFFINITY IN THE ABSENCE OF CA++  
 THAN IN ITS PRESENCE.  
 DR EMBL; M26250; G212954; -.  
 DR PIR; JQ0075; JQ0075.  
 DR PROSITE; PS00412; NEUROMODULIN\_1; 1.  
 DR PROSITE; PS00413; NEUROMODULIN\_2; 1.  
 KW NEURONE; PHOSPHORYLATION; MEMBRANE; GROWTH REGULATION;  
 KW CALMODULIN-BINDING; LIPOPROTEIN.  
 FT DOMAIN 1 4 IMPORTANT FOR MEMBRANE BINDING.  
 FT LIPID 3 3 PALMITATE (PROBABLE).  
 FT LIPID 4 4 PALMITATE (PROBABLE).  
 FT MOD\_RES 42 42 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).  
 FT DOMAIN 44 52 CALMODULIN-BINDING (PROBABLE).  
 FT DOMAIN 155 170 POLY-GLU.  
 SQ SEQUENCE 213 AA; 23569 MW; BFAC407B CRC32;  
 Query Match 68.8%; Score 44; DB 1; Length 213;  
 Best Local Similarity 55.6%; Pred. No. 1.47e+01;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Db 79 SPSEKPVE 87  
 QY 1 APPEDNPVE 9  
 RESULT 14  
 ID TRT3\_CHICK STANDARD; PRT; 262 AA.  
 AC P12620; P12621; P12619; P12618;  
 DT 01-OCT-1989 (REL. 12, CREATED)  
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE TROPONIN T, FAST SKELETAL MUSCLE ISOFORMS.  
 OS GALLUS GALLUS (CHICKEN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;  
 CC GALLIFORMES.  
 RN [1]

```

DR PROSITE; PS01132; ACTINS_ACT_LIKE; 1.
KW MULTIGENE FAMILY; STRUCTURAL PROTEIN; MUSCLE PROTEIN; ACETYLATION.
FT PROPEP 1 2
FT CHAIN 3 376 ACTIN, MUSCLE.
FT MOD_RES 3 3 ACETYLATION (BY SIMILARITY).
SQ SEQUENCE 376 AA; 41765 MW; 0A4D0295 CRC32;

Query Match 68.8%; Score 44; DB 1; Length 376;
Best Local Similarity 71.4%; Pred. No. 1 47e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 98 PPEHPV 104
    |||:|
QY 2 PPEDNPV 8

Search completed: Thu May 13 15:28:55 1999
Job time : 8 secs.

```

|        |   |
|--------|---|
| RESULT | 15  |
| ID     | ACTM-APLCA STANDARD; PRT; 376 AA.   |
| AC     | P17304;   |
| DT     | 01-AUG-1990 (REL. 15, CREATED)  |
| DT     | 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)   |
| DE     | 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)   |
| DT     | ACTIN, MUSCLE.  |
| OS     | APLYSIA CALIFORNICA (CALIFORNIA SEA HARE).  |
| OC     | EUKARYOTA; METAZOA; MOLLUSCA; GASTROPODA; OPISTHOBRANCHIA; ANASPIDEA;   |
| OC     | APLYSIIDAE.   |
| RN     | [1]   |
| RP     | SEQUENCE FROM N.A.  |
| RC     | TISSUE=MUSCLE;  |
| RX     | MEDLINE: 90301499.  |
| RA     | DESIGNERS L., AUCLAIR D., WICKHAM L.;   |
| RL     | NUCLEIC ACIDS RES. 18:3654-3654(1990).  |
| CC     | -! FUNCTION: ACTINS ARE HIGHLY CONSERVED PROTEINS THAT ARE INVOLVED IN VARIOUS TYPES OF CELL MOTILITY AND ARE UBIQUITOUSLY EXPRESSED IN ALL EUKARYOTIC CELLS.         |
| CC     | -! FUNCTION: MULTIPLE ISOFORMS ARE INVOLVED IN VARIOUS CELLULAR FUNCTIONS SUCH AS CYTOSKELETON STRUCTURE, CELL MOBILITY, CHROMOSOME MOVEMENT AND MUSCLE CONTRACTION.  |
| CC     | -! SUBUNIT: POLYMERIZATION OF GLOBULAR ACTIN (G-ACTIN) LEADS TO A STRUCTURAL FILAMENT (F-ACTIN) IN THE FORM OF A TWO-STRAINED HELIX. EACH ACTIN CAN BIND TO 4 OTHERS. |
| CC     | -! SUBCELLULAR LOCATION: CYTOPLASMIC.   |
| CC     | -! TISSUE SPECIFICITY: EXPRESSED IN THE MUSCULAR CELLS OF THE SHEATH SURROUNDING ABDOMINAL GANGLIONS.   |
| DR     | EMBL; X52866; G5568; -.   |
| DR     | PIR; S12730; S12730.  |
| DR     | HSSP; P02570; 2BTF.   |
| DR     | PROSITE; PS00406; ACTINS_1; 1.  |
| DR     | PROSITE; PS00432; ACTINS_2; 1.  |

\*\*\*\*\*  
W P S R E H  
\*\*\*\*\*  
(TW)

\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu May 13 15:28:10 1999; MasPar time 2.96 Seconds  
Tabular output not generated. 126.733 Million cell updates/sec

Title: >US-09-040-485-6  
Description: (1-10) from US09040485.pep  
Perfect Score: 64  
Sequence: 1 APPEDNPVED 10  
Scoring table: PAM 150  
Gap 15  
Searched: 116738 seqs, 37463448 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pir58  
1.pir1 2.pir2 3.pir3 4.pir4  
Statistics: Mean 21.138; Variance 27.501; scale 0.769

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No.              |
|------------|-------|-------------|--------|-------|-------------|------------------------|
| 1          | 64    | 100.0       | 757    | 2     | I38423      | aspartyl beta-hydroxy  |
| 2          | 50    | 78.1        | 704    | 1     | S60117      | protein kinase C (EC   |
| 3          | 47    | 73.4        | 129    | 2     | A70970      | hypothetical protein   |
| 4          | 46    | 71.9        | 562    | 2     | JC5182      | serine C-palmitoyltra  |
| 5          | 45    | 70.3        | 86     | 2     | S76775      | hypothetical protein   |
| 6          | 45    | 70.3        | 97     | 2     | S11755      | hypothetical protein   |
| 7          | 45    | 70.3        | 258    | 2     | C70885      | probable dehydrogenas  |
| 8          | 45    | 70.3        | 339    | 2     | B36868      | copB homolog - Xantho  |
| 9          | 45    | 70.3        | 399    | 2     | I78852      | neurofibromatosis pro  |
| 10         | 45    | 70.3        | 928    | 1     | R8HU        | retinoblastoma-associ  |
| 11         | 45    | 70.3        | 932    | 2     | F89552      | leucyl-tRNA synthetas  |
| 12         | 45    | 70.3        | 2818   | 2     | B5282       | neurofibromatosis-rel  |
| 13         | 45    | 70.3        | 2820   | 2     | JC5196      | neurofibromin I - rat  |
| 14         | 45    | 70.3        | 2825   | 2     | I54352      | neurofibromin - mouse  |
| 15         | 44    | 68.8        | 36     | 2     | A28578      | pancreatic hormone -   |
| 16         | 44    | 68.8        | 80     | 1     | PCCH        | pancreatic hormone pr  |
| 17         | 44    | 68.8        | 122    | 2     | D35392      | hypothetical protein   |
| 18         | 44    | 68.8        | 213    | 2     | QJ0075      | neuromodulin - goldfi  |
| 19         | 44    | 68.8        | 257    | 2     | B31957      | tropoinin T, skeletal  |
| 20         | 44    | 68.8        | 259    | 2     | F69311      | conserved hypothetical |
| 21         | 44    | 68.8        | 263    | 2     | C31957      | tropoinin T, skeletal  |
| 22         | 44    | 68.8        | 304    | 2     | A48174      | beta-lactamase (EC 3.  |
| 23         | 44    | 68.8        | 376    | 2     | S12730      | actin - California se  |

24 44 68.8 395 2 S36029 hypothetical protein 3.37e+01  
25 44 68.8 979 2 A35913 regulatory factor X - 3.37e+01  
26 43 67.2 176 2 S06638 calpastatin - bovine 5.20e+01  
27 43 67.2 263 2 C63044 hypothetical protein 5.20e+01  
28 43 67.2 307 2 A36885 bo-type ubiquinol oxi 5.20e+01  
29 43 67.2 491 2 A49993 glycolipptide N-tetra 5.20e+01  
30 43 67.2 506 1 FOLJG5 gag polypeptide - sim 5.20e+01  
31 43 67.2 507 2 S04237 gag polypeptide - sim 5.20e+01  
32 43 67.2 966 2 S25365 CYC8 protein - yeast 5.20e+01  
33 43 67.2 976 2 S40697 processing endoprotei 5.20e+01  
34 43 67.2 1257 2 A41060 neural cell adhesion 5.20e+01  
35 43 67.2 1259 2 S36126 neural cell adhesion 5.20e+01  
36 43 67.2 1260 2 S05479 Lepb1170\_F2\_64 protei 7.98e+01  
37 42 65.6 230 2 S72714 modulation protein no 7.98e+01  
38 42 65.6 313 2 S26838 probable DNA helicase 7.98e+01  
39 42 65.6 837 2 D71027 hypothetical protein 7.98e+01  
40 42 65.6 1133 2 S54496 probable membrane pro 7.98e+01  
41 42 65.6 1148 2 A71446 hypothetical protein 7.98e+01  
42 42 65.6 1836 2 JS0648 sodium channel alpha 7.98e+01  
43 42 65.6 1836 2 JS0648 sodium channel alpha 7.98e+01  
44 42 65.6 1836 2 JS0648 sodium channel alpha 7.98e+01  
45 42 65.6 2638 1 A42545 genome polypeptide - 7.98e+01

ALIGNMENTS

RESULT 1  
ENTRY I38423 #type complete  
TITLE aspartyl beta-hydroxylase - human  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 10-Jul-1998  
ACCESSIONS I38423  
REFERENCE I38423  
#authors Koriath, F.; Gieffers, C.; Frey, J.  
#journal Gene (1994) 150:395-399  
#title Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase.  
#cross-references MUID:95121937  
#accession I38423  
##status preliminary; translated from GB/EMBL/DBDJB  
##molecule\_type mRNA  
##residues 1-757 ##label RES  
##cross-references EMBL:U03109; NID:9458031; PID:9458032  
CLASSIFICATION #superfamily peptide-aspartate beta-dioxygenase;  
tetrairicopeptide repeat homology  
FEATURE  
54-75 #domain transmembrane #status predicted #label TRM  
SUMMARY #length 757 #molecular-weight 85498 #checksum 2143

Query Match 100.0%; Score 64; DB 2; Length 757;  
Best Local Similarity 100.0%; Pred. No. 1.75e-03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 281 APPEDNPVED 290  
Qy 1 APPEDNPVED 10

RESULT 2  
ENTRY S60117 #type complete  
TITLE protein kinase C (EC 2.7.1.-) TPA-1A - Caenorhabditis elegans  
CONTAINS protein kinase C TPA-1B  
ORGANISM #formal\_name Caenorhabditis elegans  
DATE 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 05-Sep-1997  
ACCESSIONS S60117  
REFERENCE S60117  
#authors Sano, T.; Tabuse, Y.; Nishiwaki, K.; Miwa, J.  
#journal J. Mol. Biol. (1995) 251:477-485  
#title The tpa-1 gene of Caenorhabditis elegans encodes two proteins similar to Ca(2+)-independent protein kinase Cs: evidence by complete genomic and complementary DNA sequences of the



```

tpa-1 gene.
#accession S60117
#molecule_type DNA
#residues 1-704 #label SAN
#cross-references EMBL:D49525; NID:g1217583; PID:d1009081; PID:g1217584
GENETICS
#gene tpa-1
#map_position IV
#introns 9/2; 45/1; 93/3; 128/1; 153/1; 226/3; 270/1; 446/3; 498/2; 655/3
FUNCTION
#description catalyzes the formation of peptidyl-serine-phosphate or
peptidyl-threonine-phosphate using ATP
#note activity is calcium-independent, phospholipid-dependent, and
activated by diacylglycerol and by tumor-promoting phorbol
esters
CLASSIFICATION #superfamily protein kinase C delta; protein kinase C
zinc-binding repeat homology; protein kinase homology
alternative splicing; ATP; duplication; phorbol ester
binding; phospholipid binding; phosphotransferase;
serine/threonine-specific protein kinase; zinc
FEATURE
138-704 #product protein kinase C TPA-1B #status predicted
#label MAT2\
#region pseudophosphorylation motif\
#domain protein kinase C zinc-binding repeat homology
#label K21\
#domain protein kinase C zinc-binding repeat homology
#label K22\
#domain protein kinase homology #label KIN\
#region protein kinase ATP-binding motif\
#binding_site zinc (His, Cys, Cys, Cys) #status
predicted\
179,182,204,207 #binding_site zinc (Cys, Cys, His, Cys) #status
predicted\
238,268,271,287 #binding_site zinc (His, Cys, Cys, Cys) #status
predicted\
251,254,276,279 #binding_site zinc (Cys, Cys, His, Cys) #status
predicted\
404,423,499,501 #active_site Lys, Glu, Asp, Lys #status predicted
SUMMARY #length 704 #molecular-weight 80298 #checksum 7955
Query Match 78.1%; Score 50; DB 1; Length 704;
Best Local Similarity 70.0%; Pred. No. 2.17e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 127 APDDHPVED 136
II :I:IIII
QY 1 APPEDNPVED 10
RESULT 3
ENTRY #type complete
TITLE hypothetical protein RV3354 - Mycobacterium tuberculosis
(strain H37RV)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS A70970
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry
III, C.E.; Tekata, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Squares, R.; Sulston, J.E.;
Nature (1998) 393:537-544
#journal Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.

```

```

#cross-references MUID:98295987
#accession A70970
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-129 #label COL
#cross-references GB:AL009198; GB:AL123456; NID:g3242262; PID:el202271;
PID:g2661632
#experimental_source strain H37RV
GENETICS
#gene RV3354
SUMMARY #length 129 #molecular-weight 12988 #checksum 6975
Query Match 73.4%; Score 47; DB 2; Length 129;
Best Local Similarity 60.0%; Pred. No. 8.79e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 28 APAQANPVD 37
II::IIII
QY 1 APPEDNPVED 10
RESULT 4
ENTRY #type complete
TITLE serine C-palmitoyltransferase (EC 2.3.1.50) Lcb2 chain -
Yeast (Kluyveromyces fragilis)
ALTERNATE_NAMES 3-ketosphinganine synthetase; SPT
ORGANISM #formal_name Kluyveromyces fragilis
DATE 16-Feb-1997 #sequence_revision 27-Feb-1997 #text_change
13-Sep-1998
ACCESSIONS JC5182
REFERENCE JC5180
#authors Nagiec, M.M.; Lester, R.L.; Dickson, R.C.
#journal Gene (1996) 177:237-241
#title Sphingolipid synthesis: Identification and characterization
of mammalian cDNAs encoding the Lcb2 subunit of serine
palmitoyltransferase.
#accession JC5182
#status preliminary; nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-562 #label NAG
#cross-references GB:U15646; NID:g1001948; PID:g1001949
COMMENT This enzyme catalyzes the irreversible reaction of condensation of
serine and palmitoyl-CoA to yield 3-ketosphinganine.
CLASSIFICATION #superfamily serine C-palmitoyltransferase chain LCB2
acetyltransferase
FEATURES
319-374 #region Lcb2 protein signature
SUMMARY #length 562 #molecular-weight 63004 #checksum 5581
Query Match 71.9%; Score 46; DB 2; Length 562;
Best Local Similarity 77.8%; Pred. No. 1.38e+01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 13 PREDIPLED 21
IIII I:II
QY 2 PPEDNPVED 10
RESULT 5
ENTRY #type complete
TITLE hypothetical protein - Synecocystis sp. (strain PCC 6803)
ORGANISM #formal_name Synecocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S76775
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;

```

```

Yasuda, M.; Tabata, S.
DNA Res. (1996) 3:109-136
Sequence analysis of the genome of the unicellular
cyanobacterium Synechocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUID:97061201
#accession S76775
##status preliminary
##molecule_type DNA
##residues 1-86 ##label KAN
##cross-references EMBL:D90916; GB:AB001339; NID:g1653715; PID:d1019420;
##note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
SUMMARY #length 86 #molecular-weight 9909 #checksum 7709
Query Match 70.3%; Score 45; DB 2; Length 86;
Best Local Similarity 62.5%; Pred. No. 2.17e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 49 PDDTPVEE 56
I-I I-I I-I
QY 3 PEDNPVED 10
RESULT 6
ENTRY S11755 #type fragment
TITLE hypothetical protein - Plasmodium vivax (fragment)
ORGANISM #formal_name Plasmodium vivax
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
09-Sep-1997
ACCESSIONS S11755
REFERENCE Campbell, J.R.; Franke, E.D.
#authors
#submission submitted to the EMBL Data Library, April 1989
#description A repetitive element found in Plasmodium vivax DNA.
#accession S11755
##molecule_type DNA
##residues 1-97 ##label CAM
##cross-references EMBL:X15129; NID:g10086; PID:g10087
SUMMARY #length 97 #checksum 7475
Query Match 70.3%; Score 45; DB 2; Length 97;
Best Local Similarity 71.4%; Pred. No. 2.17e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 13 PPEENPI 19
I-I I-I
QY 2 PPEDNPV 8
RESULT 7
ENTRY C70885 #type complete
TITLE probable dehydrogenase - Mycobacterium tuberculosis (strain
H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
C70885
ACCESSIONS A70500
REFERENCE Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
III, C.E.; Tekai, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsbey, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrall, B.G.
Nature (1998) 393:537-544
Deciphering the biology of Mycobacterium tuberculosis from
#journal
#title
the complete genome sequence.
#cross-references MUID:98295987
#accession C70885
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-258 ##label COL
##cross-references GB:AL008883; GB:AL123456; NID:g3261490; PID:ell172949;
PID:g2612802
##experimental_source strain H37Rv
GENETICS RV2857C
SUMMARY #length 258 #molecular-weight 26804 #checksum 1750
Query Match 70.3%; Score 45; DB 2; Length 258;
Best Local Similarity 60.0%; Pred. No. 2.17e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 93 SPEDNLIEN 102
I-I I-I I-I
QY 1 APPEDNPVED 10
RESULT 8
ENTRY B36868 #type complete
TITLE copB homolog - Xanthomonas campestris
ALTERNATE_NAMES hypothetical protein 2
ORGANISM #formal_name Xanthomonas campestris
DATE 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change
09-Sep-1997
ACCESSIONS B36868
REFERENCE Lee, Y.A.; Hendson, M.; Panopoulos, N.J.; Schroth, M.N.
J. Bacteriol. (1994) 176:173-188
#authors
#journal
#title Molecular cloning, chromosomal mapping, and sequence analysis
of copper resistance genes from Xanthomonas campestris pv.
juglandis: homology with small blue copper proteins and
multicopper oxidase.
#accession B36868
##status preliminary
##molecule_type DNA
##residues 1-339 ##label LEE
##cross-references GB:U19222; NID:g349160; PID:g461142
SUMMARY #length 339 #molecular-weight 37041 #checksum 1404
Query Match 70.3%; Score 45; DB 2; Length 339;
Best Local Similarity 60.0%; Pred. No. 2.17e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 121 PPPSDHPVHD 130
I-I I-I I-I
QY 1 APPEDNPVED 10
RESULT 9
ENTRY I78852 #type fragment
TITLE neurofibromatosis protein type 1 - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
29-Aug-1997
ACCESSIONS I78852
REFERENCE I58356
#authors Nishi, T.; Lee, P.S.; Oka, K.; Levin, V.A.; Tanase, S.;
Morino, Y.; Sava, H.
#journal Oncogene (1991) 6:1555-1559
#title Differential expression of two types of the neurofibromatosis
type 1 (NF1) gene transcripts related to neuronal
differentiation.
#cross-references MUID:92019823
#accession I78852
##status preliminary; translated from GB/EMBL/DDBJ
##molecule_type mRNA
##residues 1-399 ##label RES

```

##cross-references GB:M60915; NID:g189159; PID:g189161

## GENETICS

#gene GDB:NFI  
 ##cross-references GDB:120231; OMIM:162200  
 #map\_position 17q11.2-17q11.2  
 CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology  
 FEATURE  
 68-303 #domain ras-specific GAP catalytic domain homology  
 #label GAP  
 SUMMARY #length 399 #checksum 4004

Query Match 70.3%; Score 45; DB 2; Length 399;  
 Best Local Similarity 60.0%; Pred. No. 2.17e-01;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 380 GPEHKPVAD 389

QY 1 APPEDNPVED 10

## RESULT 10

ENTRY RBHU #type complete  
 TITLE retinoblastoma-associated protein - human  
 ALTERNATE\_NAMES retinoblastoma susceptibility protein  
 ORGANISM #formal\_name Homo sapiens #common\_name man  
 DATE 30-Jun-1987 #sequence\_revision 30-Jun-1990 #text\_change  
 05-Sep-1997  
 ACCESSIONS JS0276; A03152; A91613; A39947; A49887; I54364; I58362;  
 I78863; I78866; I78872; I78873; A35590  
 REFERENCE JS0276  
 #authors Lee, W.H.; Shew, J.Y.; Hong, F.D.; Sery, T.W.; Donoso, L.A.;  
 Young, L.J.; Bookstein, R.; Lee, E.Y.H.P.  
 #journal Nature (1987) 329:642-645  
 #title The retinoblastoma susceptibility gene encodes a nuclear  
 phosphoprotein associated with DNA binding activity.

##cross-references MUID:88014238

#accession JS0276

##molecule\_type mRNA

##residues 1-928 #label LE1  
 ##cross-references GB:M28419; NID:g190962; PID:g190963

##note this sequence has two possible initiation sites, 1-Met  
 and 113-Met

## REFERENCE

#authors A03152  
 Lee, W.H.; Bookstein, R.; Hong, F.; Young, L.J.; Shew, J.Y.;  
 Lee, E.Y.H.P.  
 #journal Science (1987) 235:1394-1399  
 #title Human retinoblastoma susceptibility gene: cloning,  
 identification, and sequence.

##cross-references MUID:87149066

#accession A03152

##molecule\_type mRNA

##residues 113-116, 'LLSYRKYT', 125-332, 'R', 334-367, 'I', 369-928  
 #label LE2

##cross-references GB:M15400; NID:g190958; PID:g190959

## REFERENCE

#authors A91613  
 McGee, T.L.; Yandell, D.W.; Dryja, T.P.  
 #journal Gene (1989) 80:119-128  
 #title Structure and partial genomic sequence of the human  
 retinoblastoma susceptibility gene.

##cross-references MUID:90006771

#accession A91613

##molecule\_type DNA

##residues 1-928 #label MCG

##cross-references GB:M27845; GB:L11910; NID:g292420; PID:g292421  
 the authors translated the codon GAA for residue 559 as  
 Gly

## REFERENCE

#authors A39947  
 Friend, S.H.; Horowitz, J.M.; Gerber, M.R.; Wang, X.F.;  
 Bogenmann, E.; Li, F.P.; Weinberg, R.A.  
 #journal Proc. Natl. Acad. Sci. U.S.A. (1987) 84:9059-9063  
 #title Deletions of a DNA sequence in retinoblastomas and  
 mesenchymal tumors: organization of the sequence and its  
 encoded protein.

##cross-references MUID:88097427

#accession A39947

##molecule\_type mRNA

##residues 1-928 #label FRI

##cross-references GB:M33647; GB:J02994; NID:g190945; PID:g190946

## REFERENCE

#authors A44987  
 T'Ang, A.; Wu, K.J.; Hashimoto, T.; Liu, W.Y.; Takahashi, R.;  
 Shi, X.H.; Mihara, K.; Zhang, F.H.; Chen, Y.Y.; Du, C.;  
 Qian, J.; Lin, Y.G.; Murphree, A.L.; Qiu, W.R.; Thompson,  
 T.; Benedict, W.F.; Fung, Y.K.T.

#journal Oncogene (1989) 4:401-407

#title Genomic organization of the human retinoblastoma gene.

##cross-references MUID:89239464

#accession A44987

##molecule\_type DNA

##residues 1-46 #label TAA

##cross-references EMBL:X16439; NID:g35894; PID:g35895

## REFERENCE

#authors I54364  
 Lohmann, D.R.; Brandt, B.; Hopping, W.; Passarge, E.;

Horsthemke, B.

#journal Hum. Mol. Genet. (1994) 3:2187-2193

#title Spectrum of small length germline mutations in the RB1 gene.

##cross-references MUID:95187159

#accession I54364

##status preliminary; translated from GB/EMBL/DBJ

##molecule\_type DNA

##residues 128-133 #label LOH

##cross-references GB:L49209; NID:g1088286; PID:g1088287

## REFERENCE

#authors I58362  
 Hogg, A.; Onadim, Z.; Baird, P.N.; Cowell, J.K.

#journal Oncogene (1992) 7:1445-1451

#title Detection of heterozygous mutations in the RB1 gene in

retinoblastoma patients using single-strand conformation

polymorphism analysis and polymerase chain reaction

sequencing.

##cross-references MUID:92319557

#accession I58362

##status preliminary; translated from GB/EMBL/DBJ

##molecule\_type DNA

##residues 1-45 #label RE2

##cross-references GB:L41889; NID:g793948; PID:g793949

#accession I78863

##status preliminary; translated from GB/EMBL/DBJ

##molecule\_type DNA

##residues 377-394 #label RE5

##cross-references GB:L41900; NID:g793969; PID:g793972

#accession I78866

##status preliminary; translated from GB/EMBL/DBJ

##molecule\_type DNA

##residues 655-671 #label RE4

##cross-references GB:L41907; NID:g801729; PID:g801730

#accession I78872

##status preliminary; translated from GB/EMBL/DBJ

##molecule\_type DNA

##residues 889-904 #label RE3

##cross-references GB:L41913; NID:g794008; PID:g794009

#accession I78873

##status preliminary; translated from GB/EMBL/DBJ

##molecule\_type DNA

##residues 906-928 #label RE3

##cross-references GB:L41914; NID:g794010; PID:g794011

## REFERENCE

#authors A35590  
 Lee, E.Y.H.P.; Bookstein, R.; Young, L.J.; Lin, C.J.;

Rosenfeld, M.G.; Lee, W.H.

#journal Proc. Natl. Acad. Sci. U.S.A. (1988) 85:6017-6021

#title Molecular mechanism of retinoblastoma gene inactivation in

retinoblastoma cell line w79.

##cross-references MUID:88320373

#contents annotation

## REFERENCE

#authors A38988  
 Lees, J.A.; Buchkovich, K.J.; Marshak, D.R.; Anderson, C.W.;

Harlow, E.

#journal EMBO J. (1991) 10:4279-4290

```

#title      The retinoblastoma protein is phosphorylated on multiple
#contents   sites by human cdc2.
#GENETICS   annotation; phosphorylation sites
#gene       GDB:RB1
##cross-references GDB:118734; OMIM:180200
#map_position 13q14.3-13q14.3
#introns    46/2; 88/3; 127/2; 167/2; 180/2; 203/1; 240/1; 287/3; 313/3;
350/2; 376/2; 405/3; 444/3; 463/3; 474/2; 500/1; 565/3;
605/2; 654/1; 702/3; 737/3; 775/3; 830/2; 840/3; 888/2;
905/1
#note       mutations in the germline gene predispose to hereditary
retinoblastoma; somatic mutations have been found in a
variety of human cancers including retinoblastoma and
osteosarcoma
CLASSIFICATION
#superfamily retinoblastoma-associated protein
KEYWORDS    cell cycle control; DNA binding; leucine zipper;
osteosarcoma; phosphoprotein; retinoblastoma; transcription
regulation; tumor suppressor
FEATURE
10-18      #region alanine-rich
20-29      #region proline-rich
662-683    #region leucine zipper motif
249,807,811 #binding_site phosphate (Ser) (covalent) (by cdc2
kinase) #status predicted
252,373    #binding_site phosphate (Thr) (covalent) (by cdc2
kinase) #status predicted
SUMMARY     #length 928 #molecular-weight 106158 #checksum 3504
Query Match 70.3%; Score 45; DB 1; Length 928;
Best Local Similarity 50.0%; Pred. NO. 2.17e-01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 27 PPEDPDQD 36
Qy 1 APPEDNPVED 10
RESULT 11
ENTRY
TITLE      F69552 #type complete
ORGANISM   leucyl-tRNA synthetase (leuS) homolog - Archaeoglobus
fulgidus
DATE       05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
21-Aug-1998
ACCESSIONS F69552
REFERENCE   A69250
#authors   Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
Peterson, J.D.; Richardson, D.L.; Kervage, A.R.; Graham,
D.E.; Kyriades, N.C.; Fleischmann, R.D.; Quackenbush, J.;
Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
T.F.; Artchach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
C.R.; Venter, J.C.
#journal   Nature (1997) 390:364-370
#title     The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MUID:98049343
#accession F69552
#status    preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues  1-932 #label KLE
##cross-references GB:AE001108; GB:AE000782; NID:g2689431; PID:g2650668;
TIGR:AR2421
CLASSIFICATION
#superfamily valine-tRNA ligase
SUMMARY     #length 932 #molecular-weight 108627 #checksum 6668

```

```

Query Match 70.3%; Score 45; DB 2; Length 932;
Best Local Similarity 66.7%; Pred. NO. 2.17e-01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 180 PHDONPVED 188
Qy 2 PPEDNPVED 10
RESULT 12
ENTRY
TITLE      B55282 #type complete
ORGANISM   neurofibromatosis-related protein NF1 - human
ALTERNATE_NAMES Gpase activating protein homolog NF1; neurofibromin
#formal_name Homo sapiens #common_name man
DATE       10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change
20-Mar-1998
ACCESSIONS B55282; A55282; A35879; A35605; A35910; A35222; A36297;
I58356
REFERENCE   A55282
#authors   Marchuk, D.A.; Saulino, A.M.; Tavakkol, R.; Swaroop, M.;
Wallace, M.R.; Andersen, L.B.; Mitchell, A.L.; Gutmann,
D.H.; Boguski, M.; Collins, F.S.
#journal   Genomics (1991) 11:931-940
#title     cDNA cloning of the type 1 neurofibromatosis gene: complete
sequence of the NF1 gene product.
#cross-references MUID:92147138
#accession B55282
#status    not compared with conceptual translation
#molecule_type mRNA
#residues  1-2818 #label MAR
##cross-references GB:M82814; NID:g189164; PID:g189165
#note      sequence extracted from NCBI backbone (NCBIP:80176)
#accession A55282
#status    preliminary
#molecule_type mRNA
#residues  1-334 #label MA2
#note      sequence extracted from NCBI backbone (NCBIN:80169,
NCBIP:80172)
REFERENCE   A35879
#authors   Xu, G.; O'Connell, P.; Viskochil, D.; Cawthon, R.; Robertson,
M.; Culver, M.; Dunn, D.; Stevens, J.; Gesteland, R.;
White, R.; Weiss, R.
#journal   Cell (1990) 62:599-608
#title     The neurofibromatosis type 1 gene encodes a protein related
to GAP.
#cross-references MUID:90335969
#accession A35879
#status    preliminary
#molecule_type mRNA
#residues  335-495, 'I', 497-1555, 'H', 1556-2818 #label XUA
##cross-references GB:M38106; GB:M57449; NID:g189169; PID:g189170
A35605
REFERENCE   A35605
#authors   Cawthon, R.M.; Weiss, R.; Xu, G.; Viskochil, D.; Culver, M.;
Stevens, J.; Robertson, M.; Dunn, D.; Gesteland, R.;
O'Connell, P.; White, R.
#journal   Cell (1990) 62:193-201
#title     A major segment of the neurofibromatosis type 1 gene: cDNA
sequence, genomic structure, and point mutations.
#cross-references MUID:90304909
#accession A35605
#status    preliminary
#molecule_type mRNA
#residues  1585-2687 #label CAW
##cross-references EMBL:M38107; EMBL:M57449
A35910
REFERENCE   A35910
#authors   Cawthon, R.M.; Weiss, R.; Xu, G.; Viskochil, D.; Culver, M.;
Stevens, J.; Robertson, M.; Dunn, D.; Gesteland, R.;
O'Connell, P.; White, R.
#journal   Cell (1990) 62:608b
#accession A35910
#status    preliminary; nucleic acid sequence not shown; not
compared with conceptual translation

```

```

##molecule_type mRNA
##residues 2688-2818 ##label CA2
REFERENCE A35222
#authors Wallace, M.R.; Marchuk, D.A.; Andersen, L.B.; Letcher, R.;
Odeh, H.M.; Saulino, A.M.; Fountain, J.W.; Brereton, A.;
Nicholson, J.; Mitchell, A.L.; Brownstein, B.H.; Collins,
F.S.
#journal Science (1990) 249:181-186
#title Type 1 neurofibromatosis gene: identification of a large
transcript disrupted in three NF1 patients.
#cross-references MIM:90319792
#accession A35222
##status preliminary
##molecule_type mRNA
##residues 2209-2818 ##label WAL
##cross-references GB:M60496; NID:g189157; PID:g189158; GB:M49193
REFERENCE A36297
#authors Martin, G.A.; Viskochil, D.; Bollag, G.; McCabe, P.C.;
Crosier, W.J.; Haubruck, H.; Conroy, L.; Clark, R.;
O'Connell, P.; Cawthon, R.M.; Innis, M.A.; McCormick, F.
#journal Cell (1990) 63:843-849
#title The GAP-related domain of the neurofibromatosis type 1 gene
product interacts with ras p21.
#cross-references MIM:91029515
#accession A36297
##molecule_type mRNA
##residues 1096-1569 'TPPEPET' ##label MA3
##cross-references GB:M61213; NID:g189162; PID:g189163
##note this clone includes an epitope tag at the 3' end
encoding the sequence TPPEPET, not part of dystrophin
but recognized by the monoclonal antibody KT3
REFERENCE I58356
#authors Nishi, T.; Lee, P.S.; Oka, K.; Levin, V.A.; Tanase, S.;
Morino, Y.; Saya, H.
#journal Oncogene (1991) 6:1555-1559
#title Differential expression of two types of the neurofibromatosis
type 1 (NF1) gene transcripts related to neuronal
differentiation.
#cross-references MIM:92019823
#accession I58356
##status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1168-1545 ##label RES
##cross-references GB:M60915; NID:g189159; PID:g189160
GENETICS
#gene GDB:NFI
#map_position 17q11.2-17q11.2
#introns 1370/3
#note the list of introns is incomplete
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
KEYWORDS alternative splicing; tumor suppressor
FEATURE
1235-1449 #domain ras-specific GAP catalytic domain homology
#label GAP
#length 2818 #molecular-weight 317030 #checksum 2858
SUMMARY
Query Match 70.3%; Score 45; DB 2; Length 2818;
Best Local Similarity 60.0%; Pred. No. 2.17e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 1526 GPPHKKPVAD 1535
:||||:|
QY 1 APPEDNPVED 10

RESULT 13
ENTRY JC5196
TITLE neurofibromin-1 - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 20-Feb-1997 #sequence_revision 27-Feb-1997 #text_change
10-Sep-1997
ACCESSIONS JC5196
#status #type complete
#molecule_type mRNA
#residues 1-2825 ##label RES
#cross-references GB:L10370; NID:g309452; PID:g309453
GENETICS
#gene NF1
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
FEATURE
1221-1456 #domain ras-specific GAP catalytic domain homology
#label GAP
#length 2825 #checksum 6076
SUMMARY
Query Match 70.3%; Score 45; DB 2; Length 2825;
Best Local Similarity 60.0%; Pred. No. 2.17e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 1533 GPPHKKPVAD 1542
:||||:|
QY 1 APPEDNPVED 10

RESULT 15
ENTRY A28578
TITLE pancreatic hormone - ostrich
ALTERNATE_NAMES pancreatic polypeptide
ORGANISM #formal_name Struthio camelus #common_name ostrich

```

```

REFERENCE JC5196
#authors Suzuki, H.; Takahashi, K.; Yasumoto, K.; Fuse, N.; Shibahara,
S.
#journal J. Biochem. (1996) 120:1048-1054
#title Differential tissue-specific expression of neurofibromin
isoform mRNAs in rat.
#accession JC5196
##status preliminary; nucleic acid sequence not shown
##molecule_type mRNA
##residues 1-2820 ##label SUZ
##cross-references DDBJ:D45201; NID:g1841313; PID:d1008732; PID:g1841314
COMMENT This protein contains a GTPase-activating protein-related domain
which is responsible for the stimulatory effect of neurofibromin
on the tyrosinase promoter activity.
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
FEATURE
1177-1436 #domain GTPase-activating protein related #status
predicted #label GRD
1237-1451 #domain ras-specific GAP catalytic domain homology
#label GAP
SUMMARY #length 2820 #molecular-weight 317080 #checksum 6628
Query Match 70.3%; Score 45; DB 2; Length 2820;
Best Local Similarity 60.0%; Pred. No. 2.17e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 1528 GPPHKKPVAD 1537
:||||:|
QY 1 APPEDNPVED 10

RESULT 14
ENTRY I54352
TITLE neurofibromin - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
25-Apr-1997
ACCESSIONS I54352
REFERENCE I54352
#authors Bernards, A.; Snijders, A.J.; Hannigan, G.E.; Murthy, A.E.;
Gusella, J.F.
#journal Hum. Mol. Genet. (1993) 2:645-650
#title Mouse neurofibromatosis type 1 cDNA sequence reveals high
degree of conservation of both coding and non-coding mRNA
segments.
#cross-references MIM:93357730
#accession I54352
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-2825 ##label RES
#cross-references GB:L10370; NID:g309452; PID:g309453
GENETICS
#gene NF1
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
FEATURE
1221-1456 #domain ras-specific GAP catalytic domain homology
#label GAP
#length 2825 #checksum 6076
SUMMARY
Query Match 70.3%; Score 45; DB 2; Length 2825;
Best Local Similarity 60.0%; Pred. No. 2.17e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 1533 GPPHKKPVAD 1542
:||||:|
QY 1 APPEDNPVED 10

RESULT 15
ENTRY A28578
TITLE pancreatic hormone - ostrich
ALTERNATE_NAMES pancreatic polypeptide
ORGANISM #formal_name Struthio camelus #common_name ostrich

```

DATE 19-Nov-1988 #sequence\_revision 19-Nov-1988 #text\_change  
 12-Apr-1995  
 A28578  
 A28578  
 #author Litthauer, D.; Oelofsen, W.  
 #journal Int. J. Pept. Protein Res. (1987) 29:739-745  
 #title Purification and primary structure of ostrich pancreatic polypeptide.  
 #cross-references MUID:87307111  
 #accession A28578  
 #molecule\_type protein  
 #residues 1-36 #label LIT  
 #note the sequence of residues 22-23 was reported as Asn-Asp in Fig. 7 and as Asp-Asn in Fig. 8  
 CLASSIFICATION #superfamily pancreatic hormone  
 SUMMARY #length 36 #molecular-weight 4209 #checksum 2998

Query Match 68.8%; Score 44; DB 2; Length 36;  
 Best Local Similarity 66.7%; Pred. NO. 3.37e+01;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 8 PGDDAPVED 16  
 | : | | | |  
 QY 2 PPEDNPVED 10

Search completed: Thu May 13 15:28:28 1999  
 Job time : 18 secs.

03 SEP 1992

|              |   |
|--------------|---|
| ACCESSIONS   | I36786  |
| REFERENCE    | A36804  |
| #authors     | Davidson, A.J.                                    |
| #submission  | submitted to GenBank, January 1992                |
| #description | Channel catfish virus: a new type of herpesvirus. |
| #accession   | I36786  |

```

##molecule_type DNA
##residues 1-232 ##label DAV
##cross-references GB:M75136; NID:g331209; PID:g331218
REFERENCE A39447
#authors Davison, A.J.
#journal Virology (1992) 186:9-14
#title Channel catfish virus: a new type of herpesvirus.
#cross-references MUID:92087490
#contents annotation
#note neither amino acid nor nucleotide sequence is given
GENETICS
#gene
CLASSIFICATION 8
#superfamily Ictalurid herpesvirus 25.5K membrane protein
transmembrane protein
FEATURE
166-182 #domain transmembrane #status predicted #label TM1\
195-211 #domain transmembrane #status predicted #label TM2
SUMMARY #length 232 #molecular-weight 25466 #checksum 8650
Query Match 74.6%; Score 50; DB 1; Length 232;
Best Local Similarity 50.0%; Pred. No. 1.10e+00;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
Db 88 ESPTGEPRD 97
QY 1 DGPTGEPOQE 10

RESULT 3
ENTRY #type complete
TITLE hypothetical protein sll1601 - Synecocystis sp. (strain PCC
6803)
ORGANISM #formal_name Synecocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S75052
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synecocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUID:97061201
#accession S75052
#status preliminary
#molecule_type DNA
#residues 1-414 ##label KAN
#cross-references EMBL:D90910; GB:AB001339; NID:g1652956; PID:d1018647;
PID:g1652997
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
SUMMARY #length 414 #molecular-weight 44810 #checksum 1664
Query Match 71.6%; Score 48; DB 2; Length 414;
Best Local Similarity 60.0%; Pred. No. 2.92e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 318 EGPTEAAQQQ 327
QY 1 DGPTGEPOQE 10

RESULT 4
ENTRY #type complete
TITLE tim (timeless) protein - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster

```

```

DATE 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change
16-Feb-1997
ACCESSIONS A57655
REFERENCE A57655
#authors Myers, M.P.; Wager-Smith, K.; Wesley, C.S.; Young, M.W.;
Sehgal, A.
#journal Science (1995) 270:805-808
#title Positional cloning and sequence analysis of the Drosophila
clock gene, timeless.
#accession A57655
#status preliminary; nucleic acid sequence not shown
#molecule_type DNA
#residues 1-1388 ##label MYE
#cross-references GB:U37018
GENETICS
#gene tim
#cross-references FlyBase:FBgn0014396
SUMMARY #length 1388 #molecular-weight 155537 #checksum 3459
Query Match 70.1%; Score 47; DB 2; Length 1388;
Best Local Similarity 60.0%; Pred. No. 4.72e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 510 DGPGKPPQHQ 519
QY 1 DGPTGEPOQE 10

RESULT 5
ENTRY #type complete
TITLE 6-deoxyerythronolide B synthase II - Saccharopolyspora
erythraea
ORGANISM #formal_name Saccharopolyspora erythraea
DATE 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change
07-Aug-1998
ACCESSIONS S23070; S22011
REFERENCE S23070
#authors Beviitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.
#journal Eur. J. Biochem. (1992) 204:39-49
#title 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora
erythraea. Cloning of the structural gene, sequence
analysis and inferred domain structure of the
multifunctional enzyme.
#accession S23070
#molecule_type DNA
#residues 1-3573 ##label BEV1
#cross-references EMBL:X62569
REFERENCE S22011
#authors Beviitt, D.J.
#submission submitted to the EMBL Data Library, September 1991
#accession S22011
#molecule_type DNA
#residues 1-184, 'I', 186-301, 'S', 303-521, 523-658, 'A', 660-993,
1001-1212, 'H', 1214-1392, 1394-2481, 'V', 2482-2827, 'P',
2829-2833, 'L', 2833-2856, 2858-2907, 'A', 2908-3135, 'K',
3137-3166, 'H', 3168-3176, 'L', 3177-3479, 'DH', 3480-3572
#label BEV2
#cross-references EMBL:X62569; NID:g46977; PID:g581651
GENETICS
#gene eryA
#start_codon GTG
CLASSIFICATION
#superfamily [acyl-carrier-protein] S-malonyltransferase
homology; 3-oxoacyl-[acyl-carrier-protein] synthase I
homology; acyl carrier protein homology; long-chain alcohol
dehydrogenase homology; short-chain alcohol dehydrogenase
homology
multifunctional enzyme
KEYWORDS
FEATURE
52-453 #domain 3-oxoacyl-[acyl-carrier-protein] synthase I
homology #label OAS1\
561-843 #domain [acyl-carrier-protein] S-malonyltransferase
homology #label AMT1\
1140-1308 #domain short-chain alcohol dehydrogenase homology

```



```
#label SAD2\
#domain acyl carrier protein homology #label ACPI\
#domain 3-oxoacyl-[acyl-carrier-protein] synthase I
homology #label OAS2\
#domain [acyl-carrier-protein] S-malonyltransferase
homology #label AMT2\
#domain long-chain alcohol dehydrogenase homology #label
LADH\
#domain short-chain alcohol dehydrogenase homology
#label SADH\
#domain acyl carrier protein homology #label ACP2
#length 3573 #molecular-weight 375258 #checksum 9583
SUMMARY
Query Match 70.1%; Score 47; DB 2; Length 3573;
Best Local Similarity 87.5%; Pred. No. 4.72e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 1384 GPTGEPAQ 1391
QY 2 GPTGEPAQ 9

RESULT 6
ENTRY #type fragment
TITLE collagen alpha 1(I) chain precursor - baboon (fragment)
ORGANISM #formal_name Papio sp. #common_name baboon
DATE 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change
21-Jul-1995
ACCESSIONS A05249
REFERENCE A92078
#authors Epstein Jr., E.H.; Scott, R.D.; Miller, E.J.; Piez, K.A.
#journal J. Biol. Chem. (1971) 246:1718-1724
#title Isolation and characterization of the peptides derived from
soluble human and baboon skin collagen after cyanogen
bromide cleavage.
#cross-references MUID:71134791
#contents CNBr0-1, CNBr2, CNBr4, CNBr5, composition
#accession A05249
#molecule_type protein
#residues 1-140 #label EPS
#experimental_source skin
COMMENT Prolines at the third position of the tripeptide repeating unit
(G-X-Y) are hydroxylated in some or all of the chains.
CLASSIFICATION #superfamily collagen alpha 1(I) chain; fibrillar collagen
carboxyl-terminal homology; von Willebrand factor type C
repeat homology
KEYWORDS coiled coil; extracellular matrix; glycoprotein; skin;
trimer; triple helix
SUMMARY #length 140 #checksum 7697
Query Match 67.2%; Score 45; DB 2; Length 140;
Best Local Similarity 44.4%; Pred. No. 1.21e+01;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 41 ZGPPGZPGZ 49
QY 1 DGPTGEPAQ 9

RESULT 7
ENTRY #type complete
TITLE Probable membrane protein ybcC - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
13-Sep-1998
ACCESSIONS G64831
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
```

```
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession G64831
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-259 #label BLAT
#cross-references GB:AE000194; GB:U00096; NID:g1787148; PID:g1787150;
UMGP:b0920
#experimental_source strain K-12, substrain MG1655
GENETICS
#gene ybcC
KEYWORDS transmembrane protein
FEATURES
13-29 #domain transmembrane #status predicted #label TM1\
39-55 #domain transmembrane #status predicted #label TM2\
SUMMARY #length 259 #molecular-weight 28666 #checksum 3266
Query Match 67.2%; Score 45; DB 2; Length 259;
Best Local Similarity 66.7%; Pred. No. 1.21e+01;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 251 GSSGEPRQE 259
QY 2 GPTGEPAQ 10

RESULT 8
ENTRY #type complete
TITLE MHC class II histocompatibility antigen B-LBII - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 03-Jun-1993 #sequence_revision 30-Sep-1993 #text_change
08-Sep-1997
ACCESSIONS A45838
REFERENCE A45838
#authors Zoorob, R.; Behar, G.; Kroemer, G.; Auffray, C.
#journal Immunogenetics (1990) 31:179-187
#title Organization of a functional chicken class II B gene.
#accession A45838
#status preliminary
#molecule_type DNA
#residues 1-263 #label ZOO
#cross-references GB:M29763; NID:g212297; PID:g212298
#note the authors translated the codon GTG for residue 112 as
Asp
CLASSIFICATION #superfamily class II histocompatibility antigen;
immunoglobulin homology
SUMMARY #length 263 #molecular-weight 29112 #checksum 5224
Query Match 67.2%; Score 45; DB 2; Length 263;
Best Local Similarity 70.0%; Pred. No. 1.21e+01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 76 DSPGEPQAE 85
QY 1 DGPTGEPAQ 10

RESULT 9
ENTRY #type fragment
TITLE P53-binding protein 1 - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change
09-Mar-1996
ACCESSIONS I38604
REFERENCE I38604
#authors Iwabuchi, K.; Bartel, P.L.; Li, B.; Marraccino, R.; Fields,
S.
#journal Proc. Natl. Acad. Sci. U.S.A. (1994) 91:6098-6102
#title Two cellular proteins that bind to wild-type but not mutant
p53.
#cross-references MUID:94286584
#accession I38604
#status preliminary; nucleic acid sequence not shown
```

```

##molecule_type mRNA
##residues 1-1027 ##label RES
##cross-references EMBL:U09477; NID:g488591; PID:g488592
SUMMARY
#length 1027 #checksum 8683

Query Match 67.2%; Score 45; DB 2; Length 1027;
Best Local Similarity 60.0%; Pred. No. 1.21e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 808 DPGTGSSEE 817
|||||:::
Qy 1 DPGTGPQOE 10

RESULT 10
ENTRY CGHU6C #type complete
TITLE Collagen alpha 1(II) chain precursor - human
ALTERNATE_NAMES collagen alpha 1(II) chain
CONTAINS chondrocalcin; collagen alpha 1(II) chain precursor splice
form 1; collagen alpha 1(II) chain precursor splice form 2;
collagen alpha 3(XI) chain
ORGANISM #formal_name Homo sapiens #common_name man
DATE 28-May-1986 #sequence_revision 01-Sep-1995 #text_change
02-Sep-1997
ACCESSIONS
A38513; S06715; S24270; A24828; S06496; A35428; A30147;
A33116; S63514; S04892; S05000; A44309; S16502; A02858;
A27280; A57033; A21733; B21733; A24561; S59491; I84433;
I37250; I37251; I37252; I37253; I37254; I55338; I59535;
I61910
REFERENCE
#authors Ryan, M.C.; Sieraski, M.; Sandell, L.J.
#journal Genomics (1990) 8:41-48
#title The human type II procollagen gene: identification of an
additional protein-coding domain and location of potential
regulatory sequences in the promoter and first intron.
#cross-references MUID:91184811
#accession A38513
##molecule_type DNA
##residues 1-103 ##label RYA
##cross-references GB:M60299; NID:g180883; PID:g18088
REFERENCE
#authors Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.
#journal Nucleic Acids Res. (1989) 17:9473
#title Nucleotide sequence of the full length cDNA encoding for
human type II procollagen.
#cross-references MUID:90067946
#accession S06715
##molecule_type mRNA
##residues 1-28,'R',99-1487 ##label SU2
##cross-references EMBL:X16468; NID:g29515; PID:g29516
#note alternative splice form 1
REFERENCE
#authors Vikkula, M.; Metsaeranta, M.; Syvaenen, A.C.; Ala-Kokko, L.;
Vuorio, E.; Peltonen, L.
#journal Biochem. J. (1992) 285:287-294
#title Structural analysis of the regulatory elements of the type-II
procollagen gene. Conservation of promoter and first intron
sequences between human and mouse.
#cross-references MUID:92344585
#accession S24270
##status translation not shown
##molecule_type DNA
##residues 1-28 ##label VIK
##cross-references EMBL:X58709; GB:S40537; NID:g35659
REFERENCE
#authors Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.
#journal Gene (1986) 44:11-16
#title Promoter region of the human pro-alpha-1-(II)-collagen gene.
#cross-references MUID:87031574
#accession A24828
##molecule_type DNA
##residues 1-8,'T',10-28 ##label NUN
##cross-references GB:M25698; NID:g180872; PID:g553237
REFERENCE
#authors Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
#journal Genomics (1989) 4:438-441
#title Organization of the exons coding for Pro alpha-1(II) collagen
N-propeptide confirms a distinct evolutionary history of
this domain of the fibrillar collagen genes.
#cross-references MUID:89233138
#accession A30147
##molecule_type DNA
##residues 104-157,'P',159-236 ##label SUM
##cross-references GB:J03065; GB:M23660; GB:M25655; GB:M25656;
GB:M25730; GB:M32168; GB:M64345; NID:g180867;
PID:g180869
REFERENCE
#authors Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
#journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:6565-6568
#title Single base mutation in the type II procollagen gene (COL2A1)
as a cause of primary osteoarthritis associated with a mild
chondrodysplasia.
#cross-references MUID:90370826
#accession A33116
##molecule_type DNA
##residues 171-172,'C',174-175 ##label ALA
#note mutant sequence from a family with family with primary
generalized osteoarthritis and mild chondrodysplasia
REFERENCE
#authors Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene
de la Porte, P.; Heibage, D.
#journal Eur. J. Biochem. (1995) 234:125-131
#title Immunohistochemical and biochemical analyses of
20000-25000-year-old fossil cartilage.
#accession S63514
##molecule_type protein
##residues 243-261;575-590;756-763,'X',765-779 ##label FRA
REFERENCE
#authors Ramirez, F.
#journal Submitted to the EMBL Data Library, December 1988
#accession S04892
##molecule_type mRNA
##residues 501-676,'A',678-783,'A',785-831,'PA',834,'F',836-1214
#note #label RAM
#cross-references EMBL:X13783; NID:g30037; PID:g930050
#accession S05000
#authors Vikkula, M.; Peltonen, L.

```

```

#journal FERS Lett. (1989) 250:171-174
#title Structural analyses of the polymorphic area in type II
#cross-references MUID:89325561
#accession S05000
##molecule_type DNA
##residues 630-640,'A',642-785 #label VIK2
##cross-references EMBL:X16158; NID:g29951
REFERENCE
A44309
#authors Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin,
D.L.; Cohn, D.H.; Eyre, D.R.
#journal J. Biol. Chem. (1992) 267:22522-22526
#title An amino acid substitution (Gly853-->Glu) in the collagen
alpha 1(I) chain produces hypochondrogenesis.
#cross-references MUID:9305458
#accession A44309
##status nucleic acid sequence not shown; not compared with
conceptual translation
##molecule_type DNA; mRNA
##residues 752-831,'PA',834,'F',836-1005,'K',1007-1036,'Q',
1038-1052,'E',1054-1068,'T',1070-1097 #label BOG
#cross-references GB:L00977; NID:g180812
##note Sequence extracted from NCBI backbone (NCBIP:117273);
parts of this sequence were determined by protein
sequencing
##note mutant sequence associated with perinatal lethal
hypochondrogenesis
REFERENCE
S16502
#authors Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
#journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:3889-3893
#title Tandem duplication within a type II collagen gene (COL2A1)
exon in an individual with spondyloepiphyseal dysplasia.
#cross-references MUID:90251662
#accession S16502
##molecule_type DNA
##residues 1164-1184,'GPSGKGANGIPGP',1185-1199 #label TIL
#cross-references EMBL:M37126; NID:g180808; PID:g180809
##note mutant sequence from a patient with spondyloepiphyseal
dysplasia
REFERENCE
A02858
#authors Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.;
Solomon, E.
#journal Proc. Natl. Acad. Sci. U.S.A. (1985) 82:2555-2559
#title Identification and characterization of the human type II
collagen gene (COL2A1).
#cross-references MUID:85190534
#accession A02858
##molecule_type DNA
##residues 1032-1056,'N',1058-1068,'T',1070-1487 #label CHE
#cross-references GB:J00115; NID:g180395; PID:g180396
REFERENCE
A27280
#authors Elima, K.; Vuorio, T.; Vuorio, E.
#journal Nucleic Acids Res. (1987) 15:9499-9504
#title Determination of the single polyadenylation site of the human
pro-alpha-1(I) collagen gene.
#cross-references MUID:88067771
#accession A27280
##molecule_type DNA; mRNA
##residues 1175-1487 #label ELI
#cross-references EMBL:X06268; NID:g30096; PID:g30097
##experimental_source fetal epiphyseal cartilage
REFERENCE
A57033
#authors van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
#journal Biochem. J. (1986) 237:923-925
#title Chondrocalcin is identical with the C-propeptide of type II
procollagen.
#accession A57033
##molecule_type protein
##residues 'XE',1244-1246,'N',1248,'X',1250-1265;1295-1305;
...
Note: remainder of annotations omitted.

```

Query Match 67.2%; Score 45; DB 1; Length 1487;

Best Local Similarity 66.7%; Pred. No. 1.21e+01;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 999 GPSGEPGQQ 1007  
||:|||||:  
Qy 2 GPTGEPQOE 10

RESULT 11  
ENTRY A33106 #type complete  
TITLE neurogenic locus mam protein - fruit fly (Drosophila melanogaster)  
ALTERNATE\_NAMES mastermind protein  
ORGANISM #formal\_name Drosophila melanogaster  
DATE 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 24-Sep-1998

ACCESSIONS A36391; A33106; S13514  
REFERENCE A36391  
#authors Smoller, D.; Friedel, C.; Schmid, A.; Bettler, D.; Lam, L.; Yedvobnick, B.  
#journal Genes Dev. (1990) 4:1688-1700  
#title The Drosophila neurogenic locus mastermind encodes a nuclear protein unusually rich in amino acid homopolymers.  
#cross-references MUID:91065516  
#accession A36391

##status preliminary  
##molecule\_type mRNA  
##residues 1-1596 #label SMO  
#cross-references GB:X54251; NID:g8203; PID:g8204  
##note strain Canton S

GENETICS  
#gene FlyBase:mam  
#cross-references FlyBase:FBgn0002643  
SUMMARY #length 1596 #molecular\_weight 167717 #checksum 4406

Query Match 67.2%; Score 45; DB 2; Length 1596;  
Best Local Similarity 66.7%; Pred. No. 1.21e+01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1317 GPMGPGPQQ 1325  
||:||||:  
Qy 2 GPTGEPQOE 10

RESULT 12  
ENTRY I54523 #type fragment  
TITLE kinesin-related protein - human (fragment)  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 05-Jun-1998  
ACCESSIONS I54523  
REFERENCE I54523  
#authors Ando, A.; Yara-Kikutl, Y.; Kawata, H.; Okamoto, N.; Imai, T.; Eki, T.; Yokoyama, K.; Soeda, E.; Ikemura, T.; Abe, K.; Inoko, H.  
#journal Immunogenetics (1994) 39:194-200  
#title Cloning of a new kinesin-related gene located at the centromeric end of the human MHC region.  
#cross-references MUID:94102819  
#accession I54523  
##status preliminary; translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-519 #label RES  
#cross-references GB:D14678; NID:g510281; PID:g510282

CLASSIFICATION #superfamily unassigned kinesin-related proteins; kinesin motor domain homology  
KEYWORDS ATP; P-loop  
FEATURE  
157-515 #domain kinesin motor domain homology #label KMOT\  
256-263 #region nucleotide-binding motif A (P-loop)  
SUMMARY #length 519 #checksum 4377

Query Match 65.7%; Score 44; DB 2; Length 519;

```

Best Local Similarity 66.7%; Pred. No. 1.92e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 269 GPGDPQLE 277
   |||:||||
QY 2 GPTGEPOE 10

RESULT 13
ENTRY S30059 #type complete
TITLE Probable transport protein ALD - human
ALTERNATE_NAMES adrenoleukodystrophy-related protein
ORGANISM #formal_name Homo sapiens #common_name man
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
24-Sep-1998
ACCESSIONS S30059
REFERENCE Mosser, J.; Douar, A.M.; Sarde, C.O.; Kioschis, P.; Feil, R.;
Moser, H.; Poustka, A.M.; Mandel, J.L.; Aubourg, P.
#journal Nature (1993) 361:726-730
#title Putative X-linked adrenoleukodystrophy gene shares unexpected
homology with ABC transporters.
#accession S30059
#molecule_type mRNA
#residues 1-745 #label MOS
#cross-references EMBL:221876; NID:g38590; PID:g38591
GENETICS
#gene GDB:ALD
#cross-references GDB:118991; OMIM:300100
#map_position Xq28-Xq28
CLASSIFICATION #superfamily ATP-binding cassette homology
KEYWORDS adrenoleukodystrophy; ATP; membrane protein; P-loop
FEATURE
490-675 #domain ATP-binding cassette homology #label ABC\
507-514 #region nucleotide-binding motif A (P-loop)\
513 #binding_site ATP (lys) #status predicted
SUMMARY #length 745 #molecular-weight 82908 #checksum 9250

Query Match 65.7%; Score 44; DB 2; Length 745;
Best Local Similarity 60.0%; Pred. No. 1.92e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 QAPAGEPTQE 56
   |||:||||
QY 1 DGTGEPOE 10

RESULT 14
ENTRY G02500 #type complete
TITLE adrenoleukodystrophy protein - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
18-Sep-1998
ACCESSIONS G02500
REFERENCE H01367
#authors Platzer, M.; Bauer, D.; Drescher, B.
#submission submitted to the EMBL Data Library, March 1996
#accession G02500
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-745 #label PLA
#cross-references EMBL:U52111; NID:g1302649; PID:g1302652
GENETICS
#gene GDB:ALD; AMN
#cross-references GDB:118991; OMIM:300100
#map_position Xq28-Xq28
#introns 300/3; 361/1; 408/3; 465/1; 496/3; 545/2; 594/1; 622/2; 664/2
CLASSIFICATION #superfamily ATP-binding cassette homology
FEATURE
490-675 #domain ATP-binding cassette homology #label ABC\
507-514 #region nucleotide-binding motif A (P-loop)
SUMMARY #length 745 #molecular-weight 82936 #checksum 9439

```

```

Query Match 65.7%; Score 44; DB 2; Length 745;
Best Local Similarity 60.0%; Pred. No. 1.92e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 QAPAGEPTQE 56
   |||:||||
QY 1 DGTGEPOE 10

RESULT 15
ENTRY S14113 #type complete
TITLE 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase (EC
3.1.4.11) delta-2 - bovine
ALTERNATE_NAMES inositol-phospholipid-specific phospholipase C
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 21-Nov-1993 #sequence_revision 02-Jun-1995 #text_change
29-May-1998
ACCESSIONS S14113; S04944
REFERENCE S14113
#authors Meldrum, E.; Kriz, R.W.; Totty, N.; Parker, P.J.
#journal Eur. J. Biochem. (1991) 196:159-165
#title A second gene product of the inositol-phospholipid-specific
phospholipase C-delta subclass.
#cross-references MUID:91160548
#accession S14113
#status preliminary
#molecule_type mRNA
#residues 1-764 #label MEL
#experimental_source brain
REFERENCE S04944
#authors Meldrum, E.; Katan, M.; Parker, P.
#journal Eur. J. Biochem. (1989) 182:673-677
#title A novel inositol-phospholipid-specific phospholipase C. Rapid
purification and characterization.
#cross-references MUID:89325315
#accession S04944
#status preliminary
#molecule_type protein
#residues 528-541, 'X', 543-553; 659-669 #label ME2
#experimental_source brain
CLASSIFICATION #superfamily 1-phosphatidylinositol-4,5-bisphosphate
phosphodiesterase III; 1-phosphatidylinositol-4,5-bisphosphate
phosphodiesterase domain X homology;
5-bisphosphate phosphodiesterase domain Y homology;
1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase
domain Y homology; pleckstrin repeat homology
phosphoric diester hydrolase
KEYWORDS
FEATURE
292-435 #domain 1-phosphatidylinositol-4,5-bisphosphate
phosphodiesterase domain X homology #label PipX\
489-609 #domain 1-phosphatidylinositol-4,5-bisphosphate
phosphodiesterase domain Y homology #label PipY
SUMMARY #length 764 #molecular-weight 87681 #checksum 1361

Query Match 65.7%; Score 44; DB 2; Length 764;
Best Local Similarity 85.7%; Pred. No. 1.92e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 340 DGPGGEP 346
   |||:||||
QY 1 DGTGEPE 7

Search completed: Thu May 13 15:36:34 1999
Job time : 12 secs.

```

\*\*\*\*\*

WATER

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu May 13 15:45:29 1999; MasPar time 2.00 Seconds  
Tabular output not generated. 133.991 Million cell updates/sec

Title: >US-09-040-485-9  
Description: (1-10) from US09040485.pep  
Perfect score: 61  
Sequence: 1 QENPDSPEPV 10

Scoring table: PAM 150  
Gap 15

Searched: 74019 seqs, 26840295 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot35  
1:swissprot

Statistics: Mean 21.064; Variance 21.843; scale 0.964

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description                       | Pred. No. |
|------------|-------|-------------|--------|----|-----------------------------------|-----------|
| 1          | 61    | 100.0       | 754    | 1  | ASPH_BOVIN ASPARTYL/ASPARAGINYL B | 1.50e-04  |
| 2          | 61    | 100.0       | 757    | 1  | ASPH_HUMAN ASPARTYL/ASPARAGINYL B | 1.50e-04  |
| 3          | 48    | 78.7        | 1379   | 1  | MET_MOUSE HEPATOCYTE GROWTH FACT  | 4.64e-01  |
| 4          | 45    | 73.8        | 304    | 1  | CBP2_SIMV1 ZINC CARBOXYPEPTIDASE  | 2.51e+00  |
| 5          | 44    | 72.1        | 662    | 1  | UL06_HSV6U VIRION PROTEIN U76     | 4.34e+00  |
| 6          | 43    | 70.5        | 277    | 1  | TRT1_HUMAN TROPONIN T, SLOW SKELE | 7.43e+00  |
| 7          | 43    | 70.5        | 395    | 1  | IT44_STRFR HYPOTHETICAL 44.4 KD P | 7.43e+00  |
| 8          | 43    | 70.5        | 475    | 1  | TRB1_ECOLI TRAB PROTEIN.          | 7.43e+00  |
| 9          | 43    | 70.5        | 995    | 1  | AGAA_VIBS7 BETA-AGARASE A PRECURS | 7.43e+00  |
| 10         | 43    | 70.5        | 1218   | 1  | YK83_YEAST PROBABLE ATP-DEPENDENT | 7.43e+00  |
| 11         | 43    | 70.5        | 1390   | 1  | MET_HUMAN HEPATOCYTE GROWTH FACT  | 7.43e+00  |
| 12         | 43    | 70.5        | 1816   | 1  | LM44_HUMAN LAMININ ALPHA-4 CHAIN  | 7.43e+00  |
| 13         | 42    | 68.9        | 622    | 1  | SKNT_YEAST PUTATIVE TRANSCRIPTION | 1.26e+01  |
| 14         | 42    | 68.9        | 1214   | 1  | TSGA_RAT TESTIS SPECIFIC PROTEI   | 1.26e+01  |
| 15         | 41    | 67.2        | 244    | 1  | SCS2_YEAST SCS2 PROTEIN.          | 2.11e+01  |
| 16         | 41    | 67.2        | 311    | 1  | SDCL_MOUSE SYNDECAN-1 PRECURSOR ( | 2.11e+01  |
| 17         | 41    | 67.2        | 313    | 1  | SDCL_RAT SYNDECAN-1 PRECURSOR (   | 2.11e+01  |
| 18         | 41    | 67.2        | 539    | 1  | PYRG_CHLTR CTP SYNTHETASE (EC 6.3 | 2.11e+01  |
| 19         | 41    | 67.2        | 642    | 1  | PHSA_STRAT PHENOXAZINONE SYNTHASE | 2.11e+01  |
| 20         | 41    | 67.2        | 662    | 1  | UL06_HSV62 VIRION PROTEIN U76.    | 2.11e+01  |
| 21         | 41    | 67.2        | 727    | 1  | KDGA_RAT DIACYLGLYCEROL KINASE,   | 2.11e+01  |
| 22         | 41    | 67.2        | 938    | 1  | EBNA_EBV EBNA-4 NUCLEAR PROTEIN   | 2.11e+01  |
| 23         | 41    | 67.2        | 1857   | 1  | FAS2_PENPA FATTY ACID SYNTHASE, S | 2.11e+01  |

## ALIGNMENTS

| RESULT | ID   | ASPH_BOVIN | STANDARD; | PRT; | 754 AA. |
|--------|--|------------|-----------|------|---------|
| AC     | Q28056;  |            |           |      |         |
| DT     | 01-NOV-1997 (REL. 35, CREATED)   |            |           |      |         |
| DT     | 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  |            |           |      |         |
| DT     | 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  |            |           |      |         |
| DE     | ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-HYDROXYLASE) (ASP BETA-HYDROXYLASE) (PEPTIDE-ASPARTATE BETA-HYDROXYLASE) |            |           |      |         |
| GN     | ASPH.  |            |           |      |         |
| OS     | BOS TAURUS (BOVINE).   |            |           |      |         |
| OC     | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;   |            |           |      |         |
| OC     | EUTHERIA; ARTIODACTYLA.  |            |           |      |         |
| RN     | [1]  |            |           |      |         |
| RP     | SEQUENCE FROM N.A.   |            |           |      |         |
| RC     | TISSUE=LIVER, AND BRAIN;   |            |           |      |         |
| RX     | MEDLINE: 92332546.   |            |           |      |         |
| RA     | JIA S., VANDUSEN W.J., DIEHL R.E., KOHL N.E., DIXON R.A.F.,  |            |           |      |         |
| RA     | ELLISTON K.O., STERN A.M., FRIEDMAN P.A.;  |            |           |      |         |
| RL     | J. BIOL. CHEM. 267:14322-14327(1992).  |            |           |      |         |
| RL     | [2]  |            |           |      |         |
| RP     | SEQUENCE OF 289-385 AND 615-641.   |            |           |      |         |
| RC     | TISSUE=LIVER;  |            |           |      |         |
| RX     | MEDLINE: 91310689.   |            |           |      |         |
| RA     | WANG O., VANDUSEN W.J., PETROSKI C.J., GARSKY V.M., STERN A.M.,  |            |           |      |         |
| RA     | FRIEDMAN P.A.;   |            |           |      |         |
| RL     | J. BIOL. CHEM. 266:14004-14010(1991).  |            |           |      |         |
| CC     | FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF PROTEINS.       |            |           |      |         |
| CC     | -1- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) -  |            |           |      |         |
| CC     | PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).   |            |           |      |         |
| CC     | -1- COFACTOR: IRON.  |            |           |      |         |
| CC     | -1- SUBUNIT: MONOMER.  |            |           |      |         |
| CC     | -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM.   |            |           |      |         |
| CC     | -1- PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 289-754) OR 52 KD (AA 311-754) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM.                   |            |           |      |         |
| DR     | EMBL: M91213; G162694.   |            |           |      |         |
| KW     | OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR; ENDOPLASMIC RETICULUM.  |            |           |      |         |
| FT     | DOMAIN 1 57  |            |           |      |         |
| FT     | TRANSMEM 58 78   |            |           |      |         |
| FT     | CYTOPLASMIC (POTENTIAL).   |            |           |      |         |
| FT     | SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).  |            |           |      |         |
| FT     | LUMENAL (POTENTIAL).   |            |           |      |         |
| FT     | POLY-GLY.  |            |           |      |         |
| FT     | POLY-SER.  |            |           |      |         |

FT DOMAIN 318 328 POLY-LYS.  
 FT CARBOHYD 96 96 POTENTIAL.  
 FT CARBOHYD 466 466 POTENTIAL.  
 FT CARBOHYD 702 702 POTENTIAL.  
 SQ SEQUENCE 754 AA; 84998 MW; 608861B2 CRC32;  
 Query Match  
 Best Local Similarity 100.0%; Score 61; DB 1; Length 754;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 248 QENPDSSEPV 257  
 QY 1 QENPDSSEPV 10  
 RESULT 2  
 ID ASPH\_HUMAN STANDARD; PRT; 757 AA.  
 AC Q12797;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DE ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-  
 DE HYDROXYLASE) (ASP BETA-HYDROXYLASE) (PEPTIDE-ASPARTATE BETA-  
 DE DIOXYGENASE).  
 GN ASPH.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 95121937.  
 RA KORIOTH F., GIEFFERS C., FREY J.;  
 RL GENE 150:393-399(1994).  
 CC -!- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN  
 CC CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF  
 CC PROTEINS.  
 CC -!- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) -  
 CC PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).  
 CC -!- COFACTOR: IRON.  
 CC -!- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC  
 CC RETICULUM.  
 CC -!- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES TESTED.  
 CC -!- PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 274-757) OR 52 KD (AA  
 CC 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY  
 CC SIMILARITY).  
 CC EMBL; U03109; E82591; .  
 DR MIM: 600582; .  
 KW OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR;  
 FT DOMAIN 1 54  
 FT TRANSMEM 55 75  
 FT CYTOPLASMIC (POTENTIAL).  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT LUMENAL (POTENTIAL).  
 FT POLY-SER.  
 FT POLY-LYS.  
 FT CARBOHYD 452 452 POTENTIAL.  
 FT CARBOHYD 705 705 POTENTIAL.  
 SQ SEQUENCE 757 AA; 85498 MW; AE6AFC24 CRC32;  
 Query Match  
 Best Local Similarity 100.0%; Score 61; DB 1; Length 757;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 233 QENPDSSEPV 242  
 QY 1 QENPDSSEPV 10  
 RESULT 3  
 ID MET\_MOUSE STANDARD; PRT; 1379 AA.  
 AC P16056; Q62125;  
 DT 01-APR-1990 (REL. 14, CREATED)

DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE HEPATOCYTE GROWTH FACTOR RECEPTOR PRECURSOR (MET PROTO-ONCOGENE  
 DE TYROSINE KINASE) (EC 2.7.1.112) (HGF-SF RECEPTOR).  
 GN MET.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 88262253.  
 RA CHAN A.M.L., KING H.W.S., DEAKIN E.A., TEMPEST P.R., HILKENS J.,  
 RA KROEZEN V., EDWARDS D.R., WILLS A.J., BROOKES P., COOPER C.S.;  
 RL ONCOGENE 2:593-599(1988).  
 RN [2]  
 RP SEQUENCE OF 1199-1270 FROM N.A.  
 RX MEDLINE: 90152381.  
 RA WILKS A.F., KURBAN R.R., HOVENS C.M., RALPH S.J.;  
 RL GENE 85:67-74(1989).  
 CC -!- FUNCTION: RECEPTOR FOR HEPATOCYTE GROWTH FACTOR. HAS A TYROSINE-  
 CC PROTEIN KINASE ACTIVITY.  
 CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -!- SUBUNIT: HETERODIMER FORMED OF AN ALPHA CHAIN (50 KD) AND A BETA  
 CC CHAIN (145 KD) WHICH ARE DISULFIDE LINKED.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -!- DISEASE: ACTIVATION OF MET AFTER REARRANGEMENT WITH THE TPR  
 CC (TRANSLOCATED PROMOTER) LOCUS OF CHROMOSOME 1 PRODUCES AN  
 CC ONCOGENIC PROTEIN.  
 CC -!- SIMILARITY: BELONGS TO THE MET TYROSINE KINASE FAMILY OF RECEPTOR.  
 CC EMBL; Y00671; G53059; .  
 DR EMBL; M33424; G200574; .  
 DR PIR; S01254; S01254.  
 DR MGI; 96969; MET.  
 DR PROSITE; PS00107; PROTEIN KINASE-ATP; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE-TYR; 1.  
 DR PROSITE; PS00111; PROTEIN KINASE-DOM; 1.  
 DR TRANSFERASE; TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; ATP-BINDING;  
 KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION; SIGNAL.  
 FT SIGNAL 1 24  
 FT CHAIN 25 1379 HEPATOCYTE GROWTH FACTOR RECEPTOR.  
 FT DOMAIN 25 931 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 932 954 POTENTIAL.  
 FT DOMAIN 955 1379 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 1076 1343 PROTEIN KINASE.  
 FT SITE 306 307 CLEAVAGE (POTENTIAL).  
 FT NP\_BIND 1082 1090 ATP (BY SIMILARITY).  
 FT BINDING 1108 1108 ATP (BY SIMILARITY).  
 FT ACT\_SITE 1202 1202 BY SIMILARITY.  
 FT MOD\_RES 1233 1233 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 45 45 POTENTIAL.  
 FT CARBOHYD 106 106 POTENTIAL.  
 FT CARBOHYD 201 201 POTENTIAL.  
 FT CARBOHYD 357 357 POTENTIAL.  
 FT CARBOHYD 398 398 POTENTIAL.  
 FT CARBOHYD 404 404 POTENTIAL.  
 FT CARBOHYD 606 606 POTENTIAL.  
 FT CARBOHYD 634 634 POTENTIAL.  
 FT CARBOHYD 784 784 POTENTIAL.  
 FT CARBOHYD 878 878 POTENTIAL.  
 FT CONFLICT 1199 1199 V -> I (IN REF. 2).  
 FT CONFLICT 1255 1255 T -> R (IN REF. 2).  
 FT CONFLICT 1261 1261 K -> T (IN REF. 2).  
 FT CONFLICT 1269 1270 VL -> ID (IN REF. 2).  
 SQ SEQUENCE 1379 AA; 153548 MW; E1597F1A CRC32;  
 Query Match  
 Best Local Similarity 78.7%; Score 48; DB 1; Length 1379;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 347 QSKPDSSEPV 356  
 QY 1 QENPDSSEPV 10

Best Local Similarity 50.0%; Pred. No. 2.51e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 17 QEHPEHVEPV 26  
QY 1 QENPDSSEPV 10

RESULT 5  
ID UL06\_HSV6U STANDARD; PRT; 662 AA.  
AC P52453;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DE VIRION PROTEIN U76.  
GN U76 OR HDLFI.  
OS HERPES SIMPLEX VIRUS (TYPE 6 / STRAIN UGANDA-1102).  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; BETAHERPESVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95027704.  
RA NICHOLAS J.;  
RL VIROLOGY 204:738-750(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95266321.  
RA GOMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,  
RA MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.;  
RL VIROLOGY 209:29-51(1995).  
CC -!- FUNCTION: PRESUMED VIRION PROTEIN; POSSIBLE ROLE IN DNA  
CC PACKAGING.  
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL6,  
CC EHV-1 56, EBV BBRF1, HCMV UL104, AND VZV 54.  
DR EMBL; U13194; G662099;  
DR EMBL; X83413; G854055;  
SQ SEQUENCE 662 AA; 77234 MW; 1F2C2F67 CRC32;

Query Match  
Best Local Similarity 72.1%; Score 44; DB 1; Length 662;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 17 QEHPEHVEPV 26  
QY 1 QENPDSSEPV 10

RESULT 4  
ID CBP2\_SIMVI STANDARD; PRT; 304 AA.  
AC P42788;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE ZINC CARBOXYPEPTIDASE (EC 3.4.17.-) (FRAGMENT).  
OS SIMULIUM VITTATUM (BLACK FLX).  
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94093864.  
RA RAMOS A., MAHOWALD A., JACOBS-LORENA M.;  
RL INSECT MOL. BIOL. 1:149-163(1993).  
CC -!- FUNCTION: INVOLVED IN THE DIGESTION OF THE BLOOD MEAL.  
CC -!- TISSUE SPECIFICITY: GUT-SPECIFIC.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE  
CC ZINC CARBOXYPEPTIDASE FAMILY.  
DR EMBL; L08481; G161186;  
DR PROSITE; PS00132; CARBOXYPEPT\_ZN\_1; 1.  
DR PROSITE; PS00133; CARBOXYPEPT\_ZN\_2; 1.  
KW HYDROLASE; CARBOXYPEPTIDASE; ZINC.  
FT NON\_TER 1  
FT METAL 58 58 ZINC (BY SIMILARITY).  
FT METAL 61 61 ZINC (BY SIMILARITY).  
FT METAL 184 184 ZINC (BY SIMILARITY).  
FT ACT\_SITE 236 236 PROTON DONOR (BY SIMILARITY).  
FT ACT\_SITE 259 259 NUCLEOPHILE (BY SIMILARITY).  
FT DISULFID 125 148 BY SIMILARITY.  
SQ SEQUENCE 304 AA; 34849 MW; 9543CCAE CRC32;

Query Match  
Best Local Similarity 73.8%; Score 45; DB 1; Length 304;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 17 QEHPEHVEPV 26  
QY 1 QENPDSSEPV 10

RESULT 5  
ID UL06\_HSV6U STANDARD; PRT; 662 AA.  
AC P52453;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DE VIRION PROTEIN U76.  
GN U76 OR HDLFI.  
OS HERPES SIMPLEX VIRUS (TYPE 6 / STRAIN UGANDA-1102).  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; BETAHERPESVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95027704.  
RA NICHOLAS J.;  
RL VIROLOGY 204:738-750(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95266321.  
RA GOMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,  
RA MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.;  
RL VIROLOGY 209:29-51(1995).  
CC -!- FUNCTION: PRESUMED VIRION PROTEIN; POSSIBLE ROLE IN DNA  
CC PACKAGING.  
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL6,  
CC EHV-1 56, EBV BBRF1, HCMV UL104, AND VZV 54.  
DR EMBL; U13194; G662099;  
DR EMBL; X83413; G854055;  
SQ SEQUENCE 662 AA; 77234 MW; 1F2C2F67 CRC32;

Query Match  
Best Local Similarity 72.1%; Score 44; DB 1; Length 662;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Best Local Similarity 50.0%; Pred. No. 4.34e+00;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 612 QTHPENSEPI 621  
QY 1 QENPDSSEPV 10

RESULT 6  
ID TTTL\_HUMAN STANDARD; PRT; 277 AA.  
AC P13805;  
DT 01-NOV-1990 (REL. 16, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE TROPONIN T, SLOW SKELETAL MUSCLE ISOFORMS.  
GN TNNI1 OR TNNI.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 88058976.  
RA GAHLMANN R., TROUTT A.B., WADE R.P., GUNNING P., KEDES L.;  
RL J. BIOL. CHEM. 262:16122-16126(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-SKELETAL MUSCLE;  
RX MEDLINE; 94183266.  
RA SAMSON F., MESNARD L., MIHOVILOVIC M., POTTER T.G., MERCADIER J.-J.,  
RA ROSES A.D., GILBERT J.R.;  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 199:841-847(1994).  
CC -!- FUNCTION: TROPONIN T IS THE TROPOMYOSIN-BINDING SUBUNIT OF  
CC TROPONIN, THE THIN FILAMENT REGULATORY COMPLEX WHICH CONFERS  
CC CALCIUM-SENSITIVITY TO STRIATED MUSCLE ACTOMYOSIN ATPASE ACTIVITY.  
CC -!- ALTERNATIVE PRODUCTS: THE DIFFERENT ISOFORMS ARE GENERATED BY A  
CC DIFFERENTIAL RNA SPLICING MECHANISM.  
DR EMBL; M19309; G339781;  
DR EMBL; M19308; G339783;  
DR EMBL; S69208; G546021;  
DR EMBL; S69209; G546023;  
DR PIR; A29783; A29783.  
DR MIM; 191041;  
KW MUSCLE PROTEIN; ALTERNATIVE SPLICING; MULTIGENE FAMILY.  
FT INIT\_MET 0  
FT MOD\_RES 1 1 PHOSPHORYLATION (BY CK2)  
FT (BY SIMILARITY).  
FT VARSPLIC 24 34 MISSING (IN SECOND ISOFORM).  
FT VARSPLIC 204 219 MISSING (IN SECOND AND THIRD ISOFORMS).  
FT CONFLICT 19 19 E -> D (IN REF. 1).  
SQ SEQUENCE 277 AA; 32817 MW; B0685CC1 CRC32;

Query Match  
Best Local Similarity 70.5%; Score 43; DB 1; Length 277;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 22 BEAPEEPEPV 31  
QY 1 QENPDSSEPV 10

RESULT 7  
ID YT44\_STRFR STANDARD; PRT; 395 AA.  
AC P20188;  
DT 01-FEB-1991 (REL. 17, CREATED)  
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 44.4 KD PROTEIN IN TRANSPOSON TN4556.  
OS STREPTOMYCES FRADIAE.  
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSON-TN4556;  
RX MEDLINE; 90185236.

```

RA SIEMIENTIAK D.R., SLIGHTOM J.L., CHUNG S.T.;
RL GENE 86:1-9(1990).
DR EMBL; M29297; G1196913; -.
PIR; JQ0430; JQ0430.
KW HYPOTHEICAL PROTEIN; TRANSPOSABLE ELEMENT.
SQ SEQUENCE 395 AA; 44379 MW; 07E0B910 CRC32;

Query Match 70.5%; Score 43; DB 1; Length 395;
Best Local Similarity 60.0%; Pred. No. 7.43e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 58 QERADGAEPV 67
|||:|:|
QY 1 QENPDSSEPV 10

RESULT 8
ID TRB1_ECOLI STANDARD; PRT; 475 AA.
AC P41067;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE TRAB PROTEIN.
GN TRAB.
OS ESCHERICHIA COLI.
OG PLASMID P. AND PLASMID COLB2.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE; 94359430.
RX FROST L.S., IPPEN-IHLER K., SKURRAY R.A.;
RL MICROBIOL. REV. 58:162-210(1994).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN-K12; PLASMID-COLB2;
RX MEDLINE; 96236035.
RA ANTHONY K.G., KATHIR P., MOORE D., IPPEN-IHLER K., FROST L.S.;
RL J. BACTERIOL. 178:3194-3200(1996).
CC -1- FUNCTION: INVOLVED IN F PILUS ASSEMBLY.
DR EMBL; U01159; G398503; -.
DR EMBL; U51860; G1293090; -.
KW PLASMID; CONJUGATION.
SQ SEQUENCE 475 AA; 50460 MW; E85059FB CRC32;

Query Match 70.5%; Score 43; DB 1; Length 475;
Best Local Similarity 50.0%; Pred. No. 7.43e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 426 EDNPEPV 435
:|:|:|
QY 1 QENPDSSEPV 10

RESULT 9
ID AGAA_VIBS7 STANDARD; PRT; 995 AA.
AC P48839;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE BETA-AGARASE A PRECURSOR (EC 3.2.1.81) (AGARASE 0107).
GN AGAA.
OS VIBRIO SP. (STRAIN JT0107).
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC VIBRIONACEAE.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE; 94113702.
RA SUGANO Y., MATSUMOTO T., KODAMA H., NOMA M.;
RL APPL. ENVIRON. MICROBIOL. 59:3750-3756(1993).
CC -1- FUNCTION: HYDROLYSES AGAROSE AND ALSO NEOGAROTETRAOSE TO YIELD
CC NEOGAROBIOSE.
CC CATALYTIC ACTIVITY: HYDROLYSIS OF 1,3-BETA-D-GALACTOSIDIC LINKAGES

```

```

CC -1- IN AGAROSE, GIVING THE TETRAMER AS THE PREDOMINANT PRODUCT.
CC -1- SIMILARITY: BELONGS TO FAMILY 50 OF GLYCOSYL HYDROLASES.
DR EMBL; D14721; G497893; -.
KW HYDROLASE; GLYCOSIDASE; SIGNAL; MULTIGENE FAMILY.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 995 BETA-AGARASE A.
FT DOMAIN 781 784 POLY-ALA.
FT DOMAIN 971 974 POLY-GLY.
SQ SEQUENCE 995 AA; 107275 MW; 9FA46198 CRC32;

Query Match 70.5%; Score 43; DB 1; Length 995;
Best Local Similarity 66.7%; Pred. No. 7.43e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 947 EAPDGEVP 955
|:|:|:|
QY 2 ENPDSSEPV 10

RESULT 10
ID YK83_YEAST STANDARD; PRT; 1218 AA.
AC P36028;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE PROBABLE ATP-DEPENDENT PERMEASE YKR103W.
GN YKR103W.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
[1]
RN SEQUENCE FROM N.A.
RP GAILLON L., DUJON B.;
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
RA -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). BELONGS TO THE MDR SUBFAMILY.
CC EMBL; Z28328; G486611; -.
DR PIR; S38182; S38182.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW HYPOTHEICAL PROTEIN; ATP-BINDING; TRANSMEMBRANE; GLYCOPROTEIN;
KW TRANSPORT.
FT TRANSMEM 136 152 POTENTIAL.
FT TRANSMEM 171 195 POTENTIAL.
FT TRANSMEM 303 327 POTENTIAL.
FT TRANSMEM 349 367 POTENTIAL.
FT TRANSMEM 455 473 POTENTIAL.
FT TRANSMEM 480 502 POTENTIAL.
FT TRANSMEM 559 582 POTENTIAL.
FT TRANSMEM 589 605 POTENTIAL.
FT DOMAIN 606 949 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 950 971 POTENTIAL.
FT TRANSMEM 1010 1034 POTENTIAL.
FT TRANSMEM 1082 1103 POTENTIAL.
FT TRANSMEM 1110 1129 POTENTIAL.
FT NP_BIND 686 693 ATP (POTENTIAL).
SQ SEQUENCE 1218 AA; 137996 MW; 42A6C96 CRC32;

Query Match 70.5%; Score 43; DB 1; Length 1218;
Best Local Similarity 77.8%; Pred. No. 7.43e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 416 ENPDSSEA 424
:|:|:|:|
QY 1 QENPDSSEPV 9

RESULT 11
ID MET_HUMAN STANDARD; PRT; 1390 AA.
AC P08581;
DT 01-AUG-1988 (REL. 08, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HEPATOCYTE GROWTH FACTOR RECEPTOR PRECURSOR (MET PROTO-ONCOGENE

```



DE TYROSINE KINASE (EC 2.7.1.112) (HGF-SF RECEPTOR).  
GN MET.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA GIORDANO S.;  
RL SUBMITTED (NOV-1990) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 87317655.  
RA PARK M., DEAN M., KAUL K., BRAUN M.J., GONDA M.A., VANDE WOUDE G.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 84:6379-6383(1987).  
RN [3]  
RP SEQUENCE OF 1010-1390 FROM N.A.  
RX MEDLINE; 88143699.  
RA CHAN A.M.L., KING H.W.S., TEMPEST P.R., DEAKIN E.A., COOPER C.S.,  
RA BROOKES P.;  
RL ONCOGENE 1:229-233(1987).  
RN [4]  
RP SEQUENCE OF 1206-1264 FROM N.A.  
RX MEDLINE; 94067791.  
RA LEE S.T., STRUNK K.M., SPRITZ R.A.;  
RL ONCOGENE 8:3403-3410(1993).  
RN [5]  
RP SEQUENCE OF 1267-1390 FROM N.A.  
RX MEDLINE; 86063462.  
RA DEAN M., PARK M., LE BEAU M.M., ROBINS T.S., DIAZ M.O., ROWLEY J.D.,  
RA BLAIR D.G., VANDE WOUDE G.F.;  
RL NATURE 318:385-388(1985).  
RN [6]  
RP SEQUENCE OF 1-754 FROM N.A.  
RA PAULEY A., ANDREWS S.;  
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [7]  
RP FUNCTION.  
RX MEDLINE; 91118019.  
RA BOTTARO D.P., RUBIN J.S., FALETTO D.L., CHAN A.M.-L., KWIETC T.E.,  
RA VANDE WOUDE G.F., AARONSON S.A.;  
RL SCIENCE 251:802-804(1991).  
RN [8]  
RP PHOSPHORYLATION AT TYR-1235.  
RX MEDLINE; 92011756.  
RA FERRACINI R., LONGATI P., NALDINI L., VIGNA E., COMOGGIO P.M.;  
RL J. BIOL. CHEM. 266:19558-19564(1991).  
RN [9]  
RP VARIANT'S HPRC, AND VARIANT VAL-320.  
RX MEDLINE; 97285124.  
RA SCHMIDT L., DUH F.-M., CHEN F., KISHIDA T., GLENN G., CHOYKE P.,  
RA SCHERER S.W., ZHUANG Z., LUBENSKY I., DEAN M., ALLIKMETS R.,  
RA CHIDAMBARAM A., BERGERHEIM U.R., FELTIS J.T., CASADEVALL C.,  
RA ZAMARRON A., BERNUES M., RICHARD S., LIPS C.J.M., WALTHER M.M.,  
RA TSUI L.-C., GEIL L., ORCUTT M.L., STACKHOUSE T., LIPAN J., SLIFE L.,  
RA BRAUCH H., DECKER J., NIEHANS G., HUGHSON M.D., MOCH H., STORKEL S.,  
RA LERMAN M.I., LINEHAN W.M., ZBAR B.;  
RL NAT. GENET. 16:68-73(1997).  
CC CC -!- FUNCTION: RECEPTOR FOR HEPATOCYTE GROWTH FACTOR. HAS A TYROSINE-  
CC CC PROTEIN KINASE ACTIVITY.  
CC CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
CC CC PROTEIN TYROSINE PHOSPHATE.  
CC CC -!- SUBUNIT: HETERODIMER FORMED OF AN ALPHA CHAIN (50 KD) AND A BETA  
CC CC CHAIN (145 KD) WHICH ARE DISULFIDE LINKED.  
CC CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC CC -!- DISEASE: ACTIVATION OF MET AFTER REARRANGEMENT WITH THE TPR  
CC CC GENE PRODUCES AN ONCOGENIC PROTEIN.  
CC CC -!- DISEASE: DEFECTS IN MET ARE THE CAUSE OF HEREDITARY PAPILLARY  
CC CC RENAL CARCINOMA (HPRC). HPRC IS A FORM OF INHERITED KIDNEY CANCER  
CC CC CHARACTERIZED BY A PREDISPOSITION TO DEVELOP MULTIPLE, BILATERAL  
CC CC PAPILLARY RENAL TUMORS. THE PATTERN OF INHERITANCE IS CONSISTENT  
CC CC WITH AUTOSOMAL DOMINANT TRANSMISSION WITH REDUCED PENETRANCE.  
CC CC -!- SIMILARITY: BELONGS TO THE MET TYROSINE KINASE FAMILY OF RECEPTOR.  
DE EMBL; M35074; G386868; -.

DR EMBL; X54559; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; J02958; G307196; -.  
DR EMBL; AC002080; G2078456; -.  
DR PIR; A40175; TVHUME.  
DR MIM; 164860; -.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
DR TRANSFERASE; TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; ATP-BINDING;  
KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION; SIGNAL;  
KW CHROMOSOMAL TRANSLOCATION; DISEASE MUTATION; POLYMORPHISM.  
FT SIGNAL 1 24  
FT CHAIN 25 1390 HEPATOCYTE GROWTH FACTOR RECEPTOR.  
FT DOMAIN 25 932 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 933 955 POTENTIAL.  
FT DOMAIN 956 1390 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 1078 1345 PROTEIN KINASE.  
FT NP\_BIND 1084 1092 ATP (BY SIMILARITY).  
FT BINDING 1110 1110 ATP (BY SIMILARITY).  
FT ACT\_SITE 1204 1204 BY SIMILARITY.  
FT SITE 307 308 CLEAVAGE (POTENTIAL).  
FT SITE 1009 1010 BREAKPOINT FOR TRANSLOCATION TO FORM  
FT TPR-MET ONCOGENE.  
FT MOD\_RES 1235 1235 PHOSPHORYLATION (AUTO-).  
FT CARBOHYD 45 45 POTENTIAL.  
FT CARBOHYD 106 106 POTENTIAL.  
FT CARBOHYD 149 149 POTENTIAL.  
FT CARBOHYD 202 202 POTENTIAL.  
FT CARBOHYD 399 399 POTENTIAL.  
FT CARBOHYD 405 405 POTENTIAL.  
FT CARBOHYD 607 607 POTENTIAL.  
FT CARBOHYD 635 635 POTENTIAL.  
FT CARBOHYD 785 785 POTENTIAL.  
FT CARBOHYD 879 879 POTENTIAL.  
FT CARBOHYD 930 930 POTENTIAL.  
FT VARIANT 320 320 A -> V.  
FT VARIANT 1131 1131 M -> L (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1188 1188 V -> L (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1195 1195 L -> V (IN HPRC; SOMATIC MUTATION).  
FT VARIANT 1220 1220 V -> I (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1228 1228 D -> N (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1228 1228 D -> H (IN HPRC; SOMATIC MUTATION).  
FT VARIANT 1230 1230 Y -> C (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1230 1230 Y -> H (IN HPRC; SOMATIC MUTATION).  
FT VARIANT 1250 1250 M -> T (IN HPRC; SOMATIC MUTATION).  
FT CONFLICT 755 755 S -> STMWKEPLNVSFLFCFAS (IN REF. 2).  
FT CONFLICT 1191 1191 G -> A (IN REF. 2).  
SQ SEQUENCE 1390 AA; 155526 MW; 650992C2 CRC32;  
Query Match 70.5%; Score 43; DB 1; Length 1390;  
Best Local Similarity 60.0%; Pred. No. 7.43e+00;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 348 QSKPDSAEPM 357  
QY 1 QENPDSSEPV 10  
| :|||:|:  
RESULT 12  
ID LMA4\_HUMAN STANDARD; PRT; 1816 AA.  
AC Q16363; Q15335; Q14735;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE LAMININ ALPHA-4 CHAIN PRECURSOR.  
GN LAMA4.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=FETAL LUNG;  
RX MEDLINE; 95300971.

RA IIVANAINEN A., SAINIO K., SARIOLA H., TRYGGVASON K.;  
 RN FEBS LETT. 365:183-188(1995).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=HEART;  
 RC RICHARDS A.J., AL-IMARA L., POPE F.M.;  
 RA SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 [3]  
 RN SEQUENCE OF 236-1816 FROM N.A.  
 RP TISSUE=HEART;  
 RC MEDLINE; 95048381.  
 RX RICHARDS A.J., AL-IMARA L., CARTER N.P., LLOYD J.C., LEVERSHA M.A.,  
 RA POPE F.M.;  
 RL GENOMICS 22:237-239(1994).  
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ  
 IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF  
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING  
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.  
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE  
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND  
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE  
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT  
 CC MEMBRANES (MAJOR COMPONENT).  
 CC -1- TISSUE SPECIFICITY: IN ADULT, STRONG EXPRESSION IN HEART, LUNG,  
 CC OVARY SMALL AND LARGE INTESTINES, PLACENTA, LIVER; WEAK OR NO  
 CC EXPRESSION IN SKELETAL MUSCLE, KIDNEY, PANCREAS, TESTIS, PROSTATE,  
 CC BRAIN. HIGH EXPRESSION IN FETAL LUNG AND KIDNEY. EXPRESSION IN  
 CC FETAL AND NEWBORN TISSUES IS OBSERVED IN CERTAIN MESENCHYMAL CELLS  
 CC IN TISSUES SUCH AS SMOOTH MUSCLE AND DERMIS.  
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT  
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.  
 CC -1- DOMAIN: DOMAIN G IS GLOBULAR.  
 CC -1- SIMILARITY: CONTAINS 3.5 LAMININ EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.  
 CC -1- CAUTION: GENE LAMA4 WAS FORMERLY CALLED LAMA3.  
 DR EMBL; S78569; G1042082;  
 DR EMBL; X91171; E198045;  
 DR EMBL; X76939; G509806;  
 DR MIM; 600133;  
 DR PROSITE; PS00222; EGF\_1; 1.  
 DR PROSITE; PS01248; LAMININ TYPE EGF; 3.  
 KW GLYCOPROTEIN; BASEMENT MEMBRANE; EXTRACELLULAR MATRIX; COILED COIL;  
 KW LAMININ EGF-LIKE DOMAIN; CELL ADHESION; REPEAT; SIGNAL.  
 FT SIGNAL 1 24  
 FT CHAIN 25 1816  
 FT DOMAIN 82 255  
 FT DOMAIN 82 131  
 FT DOMAIN 132 186  
 FT DOMAIN 187 240  
 FT DOMAIN 241 255  
 FT DOMAIN 256 851  
 FT DOMAIN 852 1816  
 FT DOMAIN 852 1027  
 FT DOMAIN 1028 1219  
 FT DOMAIN 1220 1449  
 FT DOMAIN 1450 1632  
 FT DOMAIN 1633 1816  
 FT DOMAIN 313 396  
 FT DOMAIN 466 521  
 FT DOMAIN 574 607  
 FT DOMAIN 655 717  
 FT DOMAIN 770 799  
 FT SITE 717 719  
 FT SITE 82 91  
 FT DISULFID 84 98  
 FT DISULFID 101 110  
 FT DISULFID 113 129  
 FT DISULFID 132 146  
 FT DISULFID 134 155  
 FT DISULFID 157 166  
 FT DISULFID 169 184

FT DISULFID 187 202  
 FT DISULFID 189 209  
 FT DISULFID 212 221  
 FT DISULFID 224 238  
 FT DISULFID 266 266  
 FT DISULFID 269 269  
 FT CARBOHYD 104 104  
 FT CARBOHYD 215 215  
 FT CARBOHYD 308 308  
 FT CARBOHYD 458 458  
 FT CARBOHYD 524 524  
 FT CARBOHYD 550 550  
 FT CARBOHYD 571 571  
 FT CARBOHYD 574 574  
 FT CARBOHYD 631 631  
 FT CARBOHYD 639 639  
 FT CARBOHYD 735 735  
 FT CARBOHYD 751 751  
 FT CARBOHYD 754 754  
 FT CARBOHYD 780 780  
 FT CARBOHYD 803 803  
 FT CARBOHYD 1086 1086  
 FT CARBOHYD 1281 1281  
 FT CARBOHYD 1359 1359  
 FT CARBOHYD 1411 1411  
 FT CONFLICT 143 143  
 FT CONFLICT 178 178  
 FT CONFLICT 491 491  
 FT CONFLICT 1057 1057  
 SQ SEQUENCE 1816 AA; 201964 MW; C31FEBIC CRC32;  
 Query Match 70.5%; Score 43; DB 1; Length 1816;  
 Best Local Similarity 55.6%; Pred. No. 7.43e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 45 QDPPTSEP 53  
 Qy 1 QENPDSSEP 9

RESULT 13  
 ID SKN7 YEAST STANDARD; PRT; 622 AA.  
 AC P38889; P39747;  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE PUTATIVE TRANSCRIPTION FACTOR SKN7 (POSS PROTEIN).  
 GN SKN7 OR POS9 OR YHR206W.  
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94042854.  
 RA BROWN J.L., NORTH S., BUSSEY H.;  
 RL J. BACTERIOL. 175:6908-6915(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA KREMS B., CHARIZANIS C., ENTIAN K.-D.;  
 RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / AB972;  
 RX MEDLINE; 94378003.  
 RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.;  
 RA DU Z., FAVELLO A., FULTON L., GAITUNG S., GEISEL C., KIRSTEN J.,  
 RA KUCABA T., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,  
 RA LATREILLE P., LOUIS E.J., MACRI C., MENEZES S., MOUSER L.,  
 RA NHAN M., RIFKIN L., RILES L., ST PETER H., TREVASKIS E., VAUGHAN K.,  
 RA VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,  
 RA VAUDIN M.;  
 RL SCIENCE 265:2077-2082(1994).  
 RN [4]  
 RP FUNCTION, AND MUTAGENESIS.

```

RX MEDLINE; 95045411.
RA BROWN J.L., BUSSEY H., STEWART R.C.;
RL EMBO J. 13:5186-5194(1994).
CC -!- FUNCTION: INVOLVED IN OXIDATIVE STRESS, TRANSCRIPTION FACTOR THAT
CC MAY FUNCTION IN A TWO-COMPONENT SIGNAL TRANSDUCTION PATHWAY THAT
CC ACTS IN PARALLEL WITH THE PKC1 CASCADE TO REGULATE GROWTH AT THE
CC CELL SURFACE.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE HSF FAMILY.
CC -!- SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.
DR EMBL; U00485; G414419; -
DR EMBL; X83031; G600028; -
DR EMBL; U00029; G458922; -
DR PIR; A49344; A49344.
DR PIR; S48987; S48987.
DR SGD; L0001908; SKN7.
DR PROSITE; PS00434; HSF_DOMAIN; 1.
KW TRANSCRIPTION REGULATION; SENSORY TRANSDUCTION; NUCLEAR PROTEIN;
FT DNA-BINDING; PHOSPHORYLATION.
FT DNA_BIND 86 130
FT MOD_RES 427 427 BY SIMILARITY.
FT MUTAGEN 427 427 PHOSPHORYLATION (PROBABLE).
FT MUTAGEN 427 427 D->N: DIMINISHED ACTIVITY.
FT MUTAGEN 427 427 D->E: AUGMENTED ACTIVITY.
SQ SEQUENCE 622 AA; 69202 MW; 4E506931 CRC32;

Query Match 68.9%; Score 42; DB 1; Length 622;
Best Local Similarity 50.0%; Pred. No. 1.26e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 536 QBNPSTTPV 545
QY 1 QENPDSSEPV 10

RESULT 14
ID TSGA_RAT STANDARD; PRT; 1214 AA.
AC Q63679;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TESTIS SPECIFIC PROTEIN A (ZINC FINGER PROTEIN TSGA).
GN TSGA.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE DAWLEY; TISSUE-TESTIS;
RX MEDLINE; 92172411
RA HOOG C., SCHALLING M., BRUNDELL E., DANEHOLT B.;
RL MOL. REPROD. DEV. 30:173-181(1991).
CC -!- FUNCTION: MEIOTIC OR POSTMEIOTIC FUNCTION.
CC -!- TISSUE SPECIFICITY: TESTIS-SPECIFIC; EXPRESSED ONLY IN MALE GERM
CC CELLS.
CC -!- DEVELOPMENTAL STAGE: REACHES A MAXIMUM DURING THE MEIOTIC AND THE
CC POSTMEIOTIC STAGES OF GERM CELL DEVELOPMENT.
DR EMBL; X59993; G57504; -
KW ZINC-FINGER; METAL-BINDING.
FT ZN_FING 546 571 C6-TYPE.
FT ZN_FING 546 571
SQ SEQUENCE 1214 AA; 135403 MW; 05B39332 CRC32;

Query Match 68.9%; Score 42; DB 1; Length 1214;
Best Local Similarity 50.0%; Pred. No. 1.26e+01;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 1084 QENPADHDPI 1093
QY 1 QENPDSSEPV 10

RESULT 15
ID SCS2_YEAST STANDARD; PRT; 244 AA.
AC P40075;

```

```

DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE SCS2 PROTEIN.
GN SCS2 OR YER120W.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RA NIKAWA J.-I., MURAKAMI A., ESUMI E., HOSAKA K.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,
RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,
RA CHUNG E., DUNCAN M., GUZMAN E., HARTZELL G., HUNICKE-SMITH S.,
RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,
RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGREN R., OEFNER P., OH C.,
RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOGREN T., SMITH V.,
RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: CAN SUPPRESS AN INOSITOL AUXOTROPHIC MUTANT AND A
CC CHOLINE SENSITIVE MUTANT.
DR EMBL; D44493; G624933; -
DR EMBL; U18916; G603359; -
DR SGD; L0002629; SCS2.
SQ SEQUENCE 244 AA; 26925 MW; DA1802FD CRC32;

```

Query Match 67.2%; Score 41; DB 1; Length 244;  
 Best Local Similarity 60.0%; Pred. No. 2.11e+01;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 140 QENKETVEPV 149  
 QY 1 QENPDSSEPV 10

Search completed: Thu May 13 15:45:37 1999  
 Job time : 8 secs.

(TM)

| Result No. | Query |       | Length | DB | ID     | Description            | Pred. No. |
|------------|-------|-------|--------|----|--------|------------------------|-----------|
|            | Score | Match |        |    |        |                        |           |
| 1          | 61    | 100.0 | 754    | 1  | BABOH  | peptide-aspartate bet  | 1.20e-03  |
| 2          | 61    | 100.0 | 757    | 2  | I38423 | aspartyl beta-hydroxy  | 1.20e-03  |
| 3          | 48    | 78.7  | 1379   | 2  | S01254 | hepatocyte growth fac  | 1.55e-00  |
| 4          | 46    | 75.4  | 440    | 2  | B71293 | hypothetical protein   | 4.28e-00  |
| 5          | 43    | 70.5  | 184    | 2  | A61628 | early gland protein e  | 1.87e-01  |
| 6          | 43    | 70.5  | 278    | 1  | TPHUTW | troponin T, slow skel  | 1.87e-01  |
| 7          | 43    | 70.5  | 395    | 2  | J00430 | hypothetical 44.4K pr  | 1.87e-01  |
| 8          | 43    | 70.5  | 1218   | 2  | S38182 | probable transport pr  | 1.87e-01  |
| 9          | 43    | 70.5  | 1390   | 1  | TVHUME | hepatocyte growth fac  | 1.87e-01  |
| 10         | 43    | 70.5  | 1816   | 1  | S68960 | laminin alpha-4 chain  | 1.87e-01  |
| 11         | 42    | 68.9  | 167    | 2  | E71016 | hypothetical protein   | 3.01e-01  |
| 12         | 42    | 68.9  | 607    | 2  | S27776 | 80K protein (allele c  | 3.01e-01  |
| 13         | 42    | 68.9  | 622    | 1  | A49344 | cell wall assembly re  | 3.01e-01  |
| 14         | 42    | 68.9  | 634    | 2  | JC4248 | calcium binding PW29   | 3.01e-01  |
| 15         | 42    | 68.9  | 1214   | 2  | S28499 | probable finger prote  | 3.01e-01  |
| 16         | 41    | 67.2  | 200    | 2  | H70409 | conserved hypothetical | 4.81e-01  |
| 17         | 41    | 67.2  | 244    | 2  | S50623 | SCS2 protein-1 yeast   | 4.81e-01  |
| 18         | 41    | 67.2  | 311    | 2  | S06619 | syndecan-1 precursor   | 4.81e-01  |
| 19         | 41    | 67.2  | 313    | 2  | A42853 | syndecan core protein  | 4.81e-01  |
| 20         | 41    | 67.2  | 338    | 2  | F69437 | hypothetical protein   | 4.81e-01  |
| 21         | 41    | 67.2  | 539    | 2  | B56447 | Crp synthetase homolo  | 4.81e-01  |
| 22         | 41    | 67.2  | 539    | 2  | H71545 | probable ctp syntheta  | 4.81e-01  |
| 23         | 41    | 67.2  | 727    | 2  | A56879 | diacylglycerol kinase  | 4.81e-01  |

```

KEYWORDS
FEATURE
2-56
57-78
289-754
311-754
337-370
371-404
13,96,466,702

#domain intracellular #status predicted #label INC\
#domain transmembrane #status predicted #label TRM\
#product peptide-aspartate beta-dioxygenase, 56K form
#status predicted #label 56K\
#product peptide-aspartate beta-dioxygenase, 52K form
#status predicted #label 52K\
#domain tetratricopeptide repeat homology #label TT1\
#domain tetratricopeptide repeat homology #label TT2\
#binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY
#length 754 #molecular-weight 84998 #checksum 9667

Query Match 100.0%; Score 61; DB 1: Length 754;
Best Local Similarity 100.0%; Pred. No. 1.20e-03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 248 QENPDSSEPV 257
|||||
QY 1 QENPDSSEPV 10

RESULT 2
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
AUTHORS
JOURNAL
TITLE
CROSS-REFERENCES
ACCESSION
STATUS
MOLECULE_TYPE
RESIDUES
CROSS-REFERENCES
CLASSIFICATION
FEATURE
SUMMARY

I38423 #type complete
aspartyl beta-hydroxylase - human
#formal_name Homo sapiens #common_name man
29-May-1998 #sequence_revision 29-May-1998 #text_change
10-Jul-1998
I38423
I38423
Koricho, F.; Gieffers, C.; Frey, J.
Gene (1994) 150:393-399
Cloning and characterization of the human gene encoding
aspartyl beta-hydroxylase.
#cross-references MUID:95121937
#accession I38423
#status preliminary; translated from GB/EMBL/DDBB
#molecule_type mRNA
#residues 1-757 ##label RES
##cross-references EMBL:U03109; NID:q458032
CLASSIFICATION
#superfamily peptide-aspartate beta-dioxygenase;
tetratricopeptide repeat homology

#domain transmembrane #status predicted #label TRM
#length 757 #molecular-weight 85498 #checksum 2143

Query Match 100.0%; Score 61; DB 2: Length 757;
Best Local Similarity 100.0%; Pred. No. 1.20e-03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

##residues      1-1379 ##label CHA
##Cross-references EMBL:Y00671; NID:g53058; PID:g53059
REFERENCE
JH0112
#authors      Wilks, A.F.; Kurban, R.R.; Hovens, C.M.; Ralph, S.J.
#journal      Gene (1989) 85:67-74
#title        The application of the polymerase chain reaction to cloning
              members of the protein tyrosine kinase family.
#Cross-references MUID:90152381
#accession    JH0115
#molecule_type mRNA
##residues      'I',1200-1254,'R',1256-1260,'T',1262-1268 ##label WIL
##experimental_source hemopoietic cell
##note        the authors translated the codon ACG for residue 1261 as
              Lys
REFERENCE      A45453
#authors      Weidner, K.M.; Sachs, M.; Birmmeier, W.
#journal      J. Cell Biol. (1993) 121:145-154
#title        The Met receptor tyrosine kinase transduces motility,
              proliferation, and morphogenic signals of scatter
              factor/hepatocyte growth factor in epithelial cells.
#accession    A45453
##status      preliminary; not compared with conceptual translation
##molecule_type mRNA
##residues      924-935 ##label WEI
GENETICS
#gene         met
CLASSIFICATION
#superfamily  hepatocyte growth factor receptor; protein
              kinase homology
KEYWORDS      ATP; autophosphorylation; glycoprotein; phosphoprotein;
              phosphotransferase; proto-oncogene; receptor; transmembrane
              protein; tyrosine-specific protein kinase
FEATURE
1-24          #domain signal sequence #status predicted #label SIG\
25-929        #domain extracellular #status predicted #label EXT\
25-302        #product hepatocyte growth factor receptor alpha chain
              #status predicted #label ACH\
308-1379      #product hepatocyte growth factor receptor beta chain
              #status predicted #label BCH\
930-954       #domain transmembrane #status predicted #label TMW\
955-1379      #domain intracellular #status predicted #label INT\
1074-1342     #domain protein kinase homology #label KIN\
1082-1090     #region protein kinase ATP-binding motif\
1108          #active_site Lys #status predicted\
1233         #binding_site phosphate (Tyr) (covalent) (by
              autophosphorylation) #status predicted
SUMMARY      #length 1379 #molecular_weight 153547 #checksum 6033
Query Match      78.7%; Score 48; DB 2; Length 1379;
Best Local Similarity 70.0%; Pred. No. 1.55e+00;
Matches          7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db      347 QSKPDSAEVP 356
| :|||:|
QY      1 QENPDSSEVP 10
RESULT      4
ENTRY      B71293 #type complete
TITLE      hypothetical protein TP0693 - syphilis spirochete
ORGANISM   #format_name Treponema pallidum subsp. pallidum #common_name
            syphilis spirochete
DATE       24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
            07-Aug-1998
ACCESSIONS   B71293
REFERENCE    A71250
#authors      Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.;
              Sutton, G.G.; Dodson, R.; Gwinn, M.; Hickey, E.K.; Clayton,
              R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod,
              M.P.; Salzberg, S.; Peterson, J.; Khalak, H.; Richardson,
              D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald,
              L.; Attiach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.;
              Garland, S.; Hatch, B.; Horst, K.; Roberts, K.; Watthey,
              L.; Weidman, J.; Smith, H.O.; Venter, J.C.

```



```

Tn4556
#formal_name Streptomyces fradiae
#sequence_revision 07-Sep-1990 #text_change
18-Jun-1993
ACCESSIONS
REFERENCE JQ0430
#authors Sliemantak, D.R.; Slightom, J.L.; Chung, S.T.
#journal Gene (1990) 86:1-9
#title Nucleotide sequence of Streptomyces fradiae transposable
element Tn4556: a class-II transposon related to Tn3.
#cross-references MUID:90185236
#accession JQ0430
#molecule_type DNA
#residues 1-395 #label SIE
SUMMARY #length 395 #molecular-weight 44379 #checksum 7125
Query Match 70.5%; Score 43; DB 2: Length 395;
Best Local Similarity 60.0%; Pred. No. 1.87e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 58 QBRADGAEV 67
|| :||:||||
Qy 1 QENPDSEPV 10

RESULT 8
ENTRY S38182 #type complete
TITLE Probable transport protein YKR103w - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES multidrug resistance protein homolog YKR103w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 03-May-1994 #sequence_revision 03-May-1994 #text_change
24-Jul-1998
ACCESSIONS S38182
REFERENCE S38175
#authors Gaillon, L.; Dujon, B.
#submission submitted to the Protein Sequence Database, March 1994.
#accession S38182
#molecule_type DNA
#residues 1-1218 #label GAI
#cross-references EMBL:28328; NID:9486610; PID:9486611; MIPS:YKR103w
#experimental_source strain S288C
GENETICS
#map_position 11R
CLASSIFICATION #superfamily unassigned ATP-binding cassette proteins;
ATP-binding cassette homology
ATP; P-loop; transmembrane protein
KEYWORDS
FEATURE 669-868 #domain ATP-binding cassette homology #label ABC\
686-693 #region nucleotide-binding motif A (P-loop)\
692 #binding_site ATP (Lys) #status Predicted
SUMMARY #length 1218 #molecular-weight 137995 #checksum 1891
Query Match 70.5%; Score 43; DB 2: Length 1218;
Best Local Similarity 77.8%; Pred. No. 1.87e+01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 416 EENPDSEA 424
:|||||:
Qy 1 QENPDSEPV 9

RESULT 9
ENTRY TVHUME #type complete
TITLE hepatocyte growth factor receptor precursor - human
CONTAINS protein-tyrosine kinase (EC 2.7.1.112) met
ORGANISM #formal_name Homo sapiens #common_name man
DATE 31-Mar-1991 #sequence_revision 30-Sep-1992 #text_change
20-Mar-1998
ACCESSIONS A40175; A28303; A93749; A93369; A53761; I57632; A30008;
B24569
REFERENCE A40175
#authors Giordano, S.

```

```

#submission submitted to the EMBL Data Library, November 1990
#accession A40175
#molecule_type mRNA
#residues 1-1390 #label GIO
#cross-references EMBL:X54559
REFERENCE A28303
#authors Park, M.; Dean, M.; Kaul, K.; Braun, M.J.; Gonda, M.A.; Vande
Woude, G.
#journal Proc. Natl. Acad. Sci. U.S.A. (1987) 84:6379-6383
#title Sequence of MET protooncogene cDNA has features
characteristic of the tyrosine kinase family of
growth-factor receptors.
#cross-references MUID:87317655
#accession A28303
#molecule_type mRNA
#residues 1-755, 'TWKKEPLNIVSFLFCFAS', 756-1190, 'A', 1192-1390
#cross-references GB:J02958; NID:g187558; PID:g307196
REFERENCE A93749
#authors Chan, A.M.L.; King, H.W.S.; Tempest, P.R.; Deakin, E.A.;
Cooper, C.S.; Brookes, P.
#journal Oncogene (1987) 1:229-233
#title Primary structure of the met protein tyrosine kinase domain.
#cross-references MUID:88143699
#accession A93749
#molecule_type mRNA
#residues 'VNRETCOSLRLEKLNQKALTEKNKELETAQDRNIAIOSQ',
'ETRKKELEAEKRLDIRNLSQSELYLT', 1010-1271, 'L',
1273-1390 #label CHA
#cross-references GB:U08818; NID:g487741; PID:g487742
#note this activated met oncogene is the product of gene
rearrangement
REFERENCE A93369
#authors Dean, M.; Park, M.; Le Beau, M.M.; Robins, T.S.; Diaz, M.O.;
Rowley, J.D.; Blair, D.G.; Vande Woude, G.F.
#journal Nature (1985) 318:385-388
#title The human met oncogene is related to the tyrosine kinase
oncogenes
#cross-references MUID:86065462
#accession A93369
#molecule_type DNA
#residues 1267-1390 #label DEA
#cross-references GB:M35074; NID:g187555; PID:g386868
REFERENCE A53761
#authors Gambartta, G.; Pistoi, S.; Giordano, S.; Comoglio, P.M.;
Santoro, C.
#journal J. Biol. Chem. (1994) 269:12852-12857
#title Structure and inducible regulation of the human MET promoter.
#accession A53761
#molecule_type mRNA
#residues 1-14 #label GAM
REFERENCE A40179
#authors Ferracini, R.; Longati, P.; Naldini, L.; Vigna, E.; Comoglio,
P.M.
#journal J. Biol. Chem. (1991) 266:19558-19564
#title Identification of the major autophosphorylation site of the
Met/hepatocyte growth factor receptor tyrosine kinase.
#cross-references MUID:92011756
#contents annotation; autophosphorylation site
REFERENCE I57632
#authors Dean, M.; Park, M.; Vande Woude, G.F.
#journal Mol. Cell. Biol. (1987) 7:921-924
#title Characterization of the rearranged tpr-met oncogene
breakpoint.
#cross-references MUID:87144265
#accession I57632
#status translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues 963-1009 #label RES
#cross-references GB:M15325; NID:g187531; PID:g187532
COMMENT The receptor is a dimer of disulfide-bonded 50K alpha and 145K beta
chains that arise by cleavage of the precursor. Activity is
regulated by phosphorylation of serine and tyrosine residues.

```

```

GENETICS
#gene GDB:MET
#map_position 7q31-7q31
#cross-references GDB:120178; OMIM:164860
CLASSIFICATION
#superfamily hepatocyte growth factor receptor; protein
#kinase homology
KEYWORDS
ATP; autophosphorylation; glycoprotein; phosphoprotein;
phosphotransferase; proto-oncogene; receptor; transmembrane
protein; tyrosine-specific protein kinase
FEATURE
1-24 #domain signal sequence #status predicted #label SIG\
25-303 #product hepatocyte growth factor receptor alpha chain
#status predicted #label ALP\
308-1390 #product hepatocyte growth factor receptor beta chain
#status predicted #label BET\
933-955 #domain transmembrane #status predicted #label TMN\
1076-1344 #domain protein kinase homology #label KIN\
1084-1092 #region protein kinase ATP-binding motif\
45,106,149,202,399,
405,635,785,930 #binding_site carbohydrate (Asn) (covalent) #status
predicted\
1110 #active_site lys #status experimental\
1235 #binding_site phosphate (Tyr) (covalent) (by
autophosphorylation) #status experimental
SUMMARY
#length 1390 #molecular-weight 155526 #checksum 2959
Query Match 70.5%; Score 43; DB 1; Length 1390;
Best Local Similarity 60.0%; Pred. No. 1.87e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 348 QSKPDSAEPM 357
QY 1 QENPDSSEP 10
RESULT 10
ENTRY S68960 #type complete
TITLE laminin alpha-4 chain precursor - human
AUTHOR NAMES laminin Ah
ORGANISM Homo sapiens #common_name man
DATE 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change
ACCESSIONS S68960; S65926; S49149; S40150; I53516
REFERENCE
#authors Richards, A.; Al-Imara, L.; Pope, F.M.
#journal Eur. J. Biochem. (1996) 238:813-821
#title The complete cDNA sequence of laminin alpha-4 and its
relationship to the other human laminin alpha chains.
#accession S68960
#molecule_type mRNA
#residues 1-1816 #label RIC
#cross-references EMBL:X91171; NID:g1212962; PID:s198045; PID:g1212963
#experimental_source tissue type heart
REFERENCE I53516
#authors Iivanainen, A.; Sainio, K.; Sariola, H.; Tryggvason, K.
#journal FEBS Lett. (1995) 365:183-188
#title Primary structure and expression of a novel human laminin
alpha-4 chain.
#cross-references MIM:95300971
#accession S65926
#molecule_type mRNA
#residues 1-142, 'P', 144-177, 'F', 179-490, 'Y', 492-1056, 'P', 1058-1816
#label IIV
#cross-references EMBL:S78569; NID:g1042081; PID:g1042082
REFERENCE S49149
#authors Richards, A.J.; Al-Imara, L.; Carter, N.; Leversha, M.;
Lloyd, J.C.; Pope, F.M.
#submission submitted to the EMBL Data Library, December 1993
#description Localisation of the gene (LAMA4) to chromosome 6q21 and
isolation of a partial cDNA encoding a variant laminin A
chain.
#accession S49149
#molecule_type mRNA

```

```

#residues 236-1816 #label RI2
#cross-references EMBL:X76939; NID:g509805; PID:g509806
REFERENCE S40150
#authors Richards, A.J.; Al-Imara, L.; Carter, N.; Lloyd, J.C.; Pope,
F.M.
#submission submitted to the EMBL Data Library, February 1993
#description Isolation of a partial cDNA encoding a protein homologous to
laminin A. Assignment of the gene to chromosome 6.
#accession S40150
#molecule_type mRNA
#residues 1403-1541, 'S', 1543-1816 #label RI3
#cross-references EMBL:X70904; NID:g437804; PID:g437805
GENETICS
#gene GDB:LAMA4; LAMA3
#cross-references GDB:203904; OMIM:600133
#map_position 6q21-6q21
COMPLEX
Laminins are trimers of an alpha-type, a beta-type, and a
gamma-type laminin chain.
FUNCTION
Interact with cells and with other basement membrane proteins
to promote differentiation, development, and cell migration
#superfamily laminin alpha-4 chain; laminin G repeat
homology; laminin-type EGF-like homology
KEYWORDS
basement membrane; cell binding; coiled coil; extracellular
matrix; glycoprotein; heptad repeat; heterotrimer
FEATURE
1-24 #domain signal sequence #status predicted #label SIG\
25-1816 #product laminin alpha-4 chain #status predicted #label
MAT\
82-129 #domain laminin-type EGF-like homology #label LE1\
132-184 #domain laminin-type EGF-like homology #label LE2\
187-238 #domain laminin-type EGF-like homology #label LE3\
241-265 #domain laminin-type EGF-like homology #status atypical
#label LE4\
717-719 #region cell attachment motif (R-G-D)\
871-1003 #domain laminin G repeat homology #label LG1\
1068-1198 #domain laminin G repeat homology #label LG2\
1252-1367 #domain laminin G repeat homology #label LG3\
1488-1614 #domain laminin G repeat homology #label LG4\
1665-1789 #domain laminin G repeat homology #label LG5\
104,215,308,458,
524,550,571,574,
631,639,735,751,
754,780,803,1086,
1281,1359,1411 #binding_site carbohydrate (Asn) (covalent) #status
predicted\
266,269 #disulfide_bonds interchain #status predicted
SUMMARY
#length 1816 #molecular-weight 201882 #checksum 8148
Query Match 70.5%; Score 43; DB 1; Length 1816;
Best Local Similarity 55.6%; Pred. No. 1.87e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 45 QDPETSEP 53
QY 1 QENPDSSEP 9
RESULT 11
ENTRY E71016 #type complete
TITLE hypothetical protein PH1427 - Pyrococcus horikoshii
ORGANISM Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
14-Aug-1998
REFERENCE E71016
#accessions A71000
#authors Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Oguchi,
A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.

```



```
#journal      DNA Res. (1998) 5:55-76
#title       Complete sequence and gene organization of the genome of a
             hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
             OT3.
#accession   E71016
#status      preliminary; nucleic acid sequence not shown;
             translation not shown
#molecule_type DNA
##residues   1-167 ##label KAW
##cross-references GB:AP000006; NID:g3236133; PID:di031476; PID:g3257850
##experimental_source strain OT3
##note       this accession replaces an interim accession for a
             sequence replaced by GenBank
GENETICS
#gene        PH1427
#accession   S49986
#molecule_type DNA
#length      167 #molecular-weight 18831 #checksum 3158
SUMMARY
Query Match      68.9%; Score 42; DB 2; Length 167;
Best Local Similarity 85.7%; Pred. No. 3.01e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 148 PESSEPV 154
|:|||||
QY 4 PDSSEPV 10

RESULT 12
ENTRY      S27776 #type complete
TITLE      80K protein (allele C1B) - Babesia bovis
ORGANISM   #formal_name Babesia bovis
DATE       17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
09-Sep-1997
ACCESSIONS S27776
REFERENCE   S27776
#authors   Dairymple, B.P.; Peters, J.M.
#submission submitted to the EMBL Data Library, May 1992
#description Sequence of cDNA clones of a Babesia bovis gene isolated
             using sera from cattle vaccinated with a dextran sulphate
             antigen fraction.
#accession S27776
#molecule_type mRNA
##residues 1-607 ##label DAL
##cross-references EMBL:M93126; NID:g155862; PID:g155863
SUMMARY    #length 607 #molecular-weight 67129 #checksum 8353
Query Match      68.9%; Score 42; DB 2; Length 607;
Best Local Similarity 55.6%; Pred. No. 3.01e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 367 EEPEAPEPV 375
|:|::|||
QY 2 ENPDSSEPV 10

RESULT 13
ENTRY      A49344 #type complete
TITLE      cell wall assembly regulatory protein SKN7 - yeast
             (Saccharomyces cerevisiae)
ALTERNATE_NAMES
ALTERNATE_NAMES oxidative stress response regulator POS9; protein YHR206w
ORGANISM   #formal_name Saccharomyces cerevisiae
DATE       07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change
17-Apr-1998
ACCESSIONS A49344; S48987; S49986; S68114
REFERENCE   A49344
#authors   Brown, J.L.; North, S.; Bussey, H.
#journal   J. Bacteriol. (1993) 175:6908-6915
#title     SKN7, a yeast multicopy suppressor of a mutation affecting
             cell wall beta-glucan assembly, encodes a product with
             domains homologous to prokaryotic two-component regulators
             and to heat shock transcription factors.
#accession A49344
#molecule_type DNA
##residues 1-622 ##label BRO
```

```
#cross-references GB:U00485; NID:g414418; PID:g414419
REFERENCE   S46671
#authors   Macri, C.
#submission submitted to the EMBL Data Library, February 1994
#description The sequence of S. cerevisiae cosmid 9177.
#accession S48987
#molecule_type DNA
##residues 1-622 ##label MAC
##cross-references EMBL:U00029; NID:g551322; PID:g458922; MIPS:YHR206w
REFERENCE   S49986
#authors   Krens, B.; Charizanis, C.; Entian, K.D.
#submission submitted to the EMBL Data Library, November 1994
#description A protein (Pos9) similar to prokaryotic response regulators
             is involved in oxidative stress in yeast.
#accession S49986
#molecule_type DNA
##residues 1-622 ##label KRE
##cross-references EMBL:X83031; NID:g600027; PID:g600028
REFERENCE   S68114
#authors   Krens, B.; Charizanis, C.; Entian, K.D.
#journal   Curr. Genet. (1996) 29:327-334
#title     The response regulator-like protein Pos9/Skn7 of
             Saccharomyces cerevisiae is involved in oxidative stress
             resistance.
#accession S68114
#status    nucleic acid sequence not shown; translation not shown
#molecule_type DNA
##residues 1-622 ##label KRW
##cross-references EMBL:X83031; NID:g600027; PID:g600028
##note     the nucleotide sequence was submitted to the EMBL Data
             Library, November 1994
GENETICS
#gene      SGD:SKN7; POS9
#accession SGD:S0001249; MIPS:YHR206w
#map_position 8R
CLASSIFICATION #superfamily cell wall assembly regulatory protein SKN7; HSF
               DNA-binding domain homology; response regulator homology
KEYWORDS       DNA binding; leucine zipper; nucleus; phosphoprotein;
               transcription regulation
FEATURE
87-194      #domain HSF DNA-binding domain homology #label HSF\
379-488     #domain response regulator homology #label RRH\
382-410     #region leucine zipper\
555-576     #region glutamine-rich\
427         #binding-site phosphate (Asp) (covalent) #status
               predicted
SUMMARY       #length 622 #molecular-weight 69202 #checksum 256
Query Match    68.9%; Score 42; DB 1; Length 622;
Best Local Similarity 50.0%; Pred. No. 3.01e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 536 QDNPSTTPV 545
|:|::|||
QY 1 QENPDSSEPV 10

RESULT 14
ENTRY      JC4248 #type complete
TITLE      calcium binding PW29 protein - mouse
ORGANISM   #formal_name Mus musculus #common_name house mouse
DATE       12-Oct-1995 #sequence_revision 08-Feb-1996 #text_change
07-Nov-1997
ACCESSIONS JC4248
REFERENCE   JC4248
#authors   Yu, S.; Ozawa, M.; Naved, A.F.; Miyauchi, T.; Muramatsu, H.;
             Muramatsu, T.
#journal   Cell Struct. Funct. (1995) 20:263-268
#title     cDNA cloning and sequence analysis of a novel calcium binding
             protein with oligoproline motif.
#accession JC4248
#molecule_type mRNA
##residues 1-634 ##label YUS
```

##cross-references DBJ:D49429; NID:g599609; PID:dl009004; PID:gl304155  
##experimental\_source F9 embryonal carcinoma cells

COMMENT This protein is a cytoplasmic calcium binding protein which lacks EF-hand motif, and is present in embryonal carcinoma cells. It plays important roles in regulation of cellular activities. This protein is rich in hydrophilic amino acids.

## KEYWORDS

calcium binding

## FEATURE

528-547 #region glutamic acid/lysine-rich  
SUMMARY #length 634 #molecular-weight 71893 #checksum 7243

## Query Match

68.9%; Score 42; DB 2; Length 634;

Best Local Similarity 55.6%; Pred. No. 3.01e+01;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 276 DSPDSVDPV 284

::|||::||

QY 2 ENPDSSEPV 10

## RESULT 15

## ENTRY

S28499 #type complete

probable finger protein - rat

TITLE #formal\_name Rattus norvegicus #common\_name Norway rat

ORGANISM 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change

DATE 10-Sep-1997

## ACCESSIONS

S28499

## REFERENCE

S28499

#authors Hoog, C.; Schalling, M.; Grunder-Brundell, E.; Daneholt, B.

#submission submitted to the EMBL Data Library, June 1991

#description Analysis of a murine germ cell-specific transcript that

#accession encodes a putative zinc finger protein.

#molecule\_type mRNA

#residues 1-1214 #label HOO

#cross-references EMBL:X59993; NID:g57503; PID:g57504

#experimental\_source strain Sprague-Dawley

KEYWORDS DNA binding; zinc; zinc finger

SUMMARY #length 1214 #molecular-weight 135403 #checksum 4667

## Query Match

68.9%; Score 42; DB 2; Length 1214;

Best Local Similarity 50.0%; Pred. No. 3.01e+01;

Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 1084 QENPADHDPI 1093

||||::||

QY 1 QENPDSSEPV 10

Search completed: Thu May 13 15:45:09 1999

Job time : 17 secs.

\*\*\*\*\*  
M P S R L A  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 13 15:45:58 1999; MasPar time 4.01 Seconds  
Tabular output not generated. 124.082 Million cell updates/sec

Title: >US-09-040-485-9  
Description: (1-10) from US09040485.pep  
Perfect Score: 61  
Sequence: 1 QENPDSPEPV 10

Scoring table: PAM 150  
Gap 15

Searched: 165420 seqs, 49795644 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrnbl6  
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human  
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle  
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified  
13:sp.vertebrate 14:sp.virus

Statistics: Mean 20.152; Variance 22.779; scale 0.885

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                     | Pred. No. |
|------------|-------|-------------|--------|-------|---------------------------------|-----------|
| 1          | 49    | 80.3        | 374    | 13    | 042322 D4B DOPAMINE RECEPTOR.   | 6.39e+01  |
| 2          | 46    | 75.4        | 344    | 5     | 062447 Y43F4B.1.                | 3.26e+00  |
| 3          | 44    | 72.1        | 590    | 6     | 002665 SODIUM-D-GLUCOSE COTRA   | 9.31e+00  |
| 4          | 44    | 72.1        | 628    | 2     | 067990 METALLOPROTEINASE (FRAGM | 9.31e+00  |
| 5          | 44    | 72.1        | 1262   | 5     | 020771 F54D5.5.                 | 9.31e+00  |
| 6          | 43    | 70.5        | 105    | 4     | 014421 GLYCOPHORIN M2 (FRAGME   | 1.56e+01  |
| 7          | 43    | 70.5        | 129    | 4     | 099737 LAMININ ALPHA 4 CHAIN.   | 1.56e+01  |
| 8          | 43    | 70.5        | 158    | 5     | 020474 F46F6.3.                 | 1.56e+01  |
| 9          | 43    | 70.5        | 184    | 5     | 024743 EGP-1 PRECURSOR.         | 1.56e+01  |
| 10         | 43    | 70.5        | 480    | 5     | 027033 MEMBRANE PROTEIN.        | 1.56e+01  |
| 11         | 43    | 70.5        | 1382   | 11    | 097523 HGF RECEPTOR PRECURSOR   | 1.56e+01  |
| 12         | 43    | 70.5        | 1382   | 11    | 097523 HEPATOCYTE GROWTH FACT   | 1.56e+01  |
| 13         | 42    | 68.9        | 81     | 4     | 099463 GLYCOSTYLTRANSFERASE (F  | 2.58e+01  |
| 14         | 42    | 68.9        | 139    | 5     | 091281 COSMID F27C1.            | 2.58e+01  |
| 15         | 42    | 68.9        | 156    | 10    | 039754 GRPF1.                   | 2.58e+01  |
| 16         | 42    | 68.9        | 167    | 1     | 059097 167AA LONG HYPOTHETICA   | 2.58e+01  |
| 17         | 42    | 68.9        | 345    | 14    | 056987 COAT PROTEIN.            | 2.58e+01  |
| 18         | 42    | 68.9        | 607    | 5     | 017112 80 KDA PROTEIN.          | 2.58e+01  |
| 19         | 42    | 68.9        | 631    | 4     | 099568 KIAA0078 PROTEIN.        | 2.58e+01  |
| 20         | 42    | 68.9        | 631    | 4     | 060216 PROTEIN INVOLVED IN DN   | 2.58e+01  |

|    |    |      |      |    |                               |          |
|----|----|------|------|----|-------------------------------|----------|
| 21 | 42 | 68.9 | 634  | 11 | Q61550 RAD21 HOMOLOG (S. POMB | 2.58e+01 |
| 22 | 42 | 68.9 | 635  | 11 | P70219 RAD21 HOMOLOG (S.      | 2.58e+01 |
| 23 | 42 | 68.9 | 1211 | 11 | O35233 MAF4 (FRAGMENT).       | 2.58e+01 |
| 24 | 41 | 67.2 | 85   | 5  | Q23300 C01F6.8.               | 4.24e+01 |
| 25 | 41 | 67.2 | 140  | 2  | Q45022 XYLX GENE (FRAGMENT).  | 4.24e+01 |
| 26 | 41 | 67.2 | 157  | 6  | O62841 CATHALICIDIN PRECURSOR | 4.24e+01 |
| 27 | 41 | 67.2 | 200  | 2  | O67307 HYPOTHETICAL 23.1 KD P | 4.24e+01 |
| 28 | 41 | 67.2 | 211  | 2  | O07446 INSERTION ELEMENT IS14 | 4.24e+01 |
| 29 | 41 | 67.2 | 283  | 5  | O26940 HYPOTHETICAL P284 PROT | 4.24e+01 |
| 30 | 41 | 67.2 | 338  | 1  | O28769 HYPOTHETICAL 37.4 KD P | 4.24e+01 |
| 31 | 41 | 67.2 | 405  | 10 | O22763 PUTATIVE BZIP-LIKE DNA | 4.24e+01 |
| 32 | 41 | 67.2 | 482  | 2  | O52762 CATALASE (EC 1.11.1.6) | 4.24e+01 |
| 33 | 41 | 67.2 | 484  | 2  | P77939 CATALASE (EC 1.11.1.6) | 4.24e+01 |
| 34 | 41 | 67.2 | 493  | 9  | O64203 GP10.                  | 4.24e+01 |
| 35 | 41 | 67.2 | 625  | 11 | O35305 RANK.                  | 4.24e+01 |
| 36 | 41 | 67.2 | 662  | 14 | O57138 SIMILAR TO HHV6A U76.  | 4.24e+01 |
| 37 | 41 | 67.2 | 919  | 10 | O24375 ALPHA-GLUCOSIDASE (EC  | 4.24e+01 |
| 38 | 41 | 67.2 | 946  | 14 | Q69139 NUCLEAR ANTIGEN EBNA-3 | 4.24e+01 |
| 39 | 41 | 67.2 | 991  | 5  | O18152 T28D6.4.               | 4.24e+01 |
| 40 | 41 | 67.2 | 1186 | 14 | O90061 NUCLEAR ANTIGEN-3B (EX | 4.24e+01 |
| 41 | 41 | 67.2 | 1555 | 5  | O44498 F15E6.1 PROTEIN.       | 4.24e+01 |
| 42 | 40 | 65.6 | 330  | 11 | O54950 AMP ACTIVATED PROTEIN  | 6.90e+01 |
| 43 | 40 | 65.6 | 586  | 2  | P96351 CTP SYNTHETASE (EC 6.3 | 6.90e+01 |
| 44 | 40 | 65.6 | 702  | 1  | O28289 HYPOTHETICAL 78.9 KD P | 6.90e+01 |
| 45 | 40 | 65.6 | 1078 | 11 | O08995 MYELIN TRANSCRIPTION F | 6.90e+01 |

ALIGNMENTS

|   |   |      |         |
|---|---|------|---------|
| RESULT 1  | PRELIMINARY;                                      | PRT; | 374 AA. |
| ID 042322   | 01-JAN-1998 (TREMREL. 05, CREATED)                |      |         |
| AC 042322   | 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)   |      |         |
| DT 01-JAN-1998  | 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE) |      |         |
| DE D4B DOPAMINE RECEPTOR.   |   |      |         |
| OS CYPRINUS CARPIO (COMMON CARP).                                   |   |      |         |
| OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNATHOSTOMATA; |   |      |         |
| OC OSTEICHTHYES; ACTINOPTERYGII; CYPRINIFORMES.                     |   |      |         |
| RN [1]  |   |      |         |
| RP SEQUENCE FROM N.A.   |   |      |         |
| RC TISSUE=RETINA;   |   |      |         |
| RA HIRANO J., ARCHER S.N., DJAMGOZ M.B.A.;                          |   |      |         |
| RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.            |   |      |         |
| DR EMBL; Y14633; E334823; -   |   |      |         |
| DR PFAM; PF00001; 7tm_1.  |   |      |         |
| SQ SEQUENCE 374 AA; 42004 MW; 185EF905 CRC32;                       |   |      |         |

Query Match 80.3%; Score 49; DB 13; Length 374;  
Best Local Similarity 60.0%; Pred. No. 6.39e+01;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

|                      |              |      |         |
|----------------------|--------------|------|---------|
| Db 255 EQDPSPEPV 264 | PRELIMINARY; | PRT; | 344 AA. |
| Qy 1 QENPDSPEPV 10   |              |      |         |

|   |   |      |         |
|---|---|------|---------|
| RESULT 2  | PRELIMINARY;                                      | PRT; | 344 AA. |
| ID 062447   | 01-AUG-1998 (TREMREL. 07, CREATED)                |      |         |
| AC 062447   | 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)   |      |         |
| DT 01-AUG-1998  | 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE) |      |         |
| DE Y43F4B.1   |   |      |         |
| OS CAENORHABDITIS ELEGANS.  |   |      |         |
| OC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTIA; RHABDITIDA.    |   |      |         |
| RN [1]  |   |      |         |
| RP SEQUENCE FROM N.A.   |   |      |         |
| RA MATTHEWS L.;   |   |      |         |
| RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.                  |   |      |         |
| RN [2]  |   |      |         |
| RP SEQUENCE FROM N.A.   |   |      |         |
| RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J., |   |      |         |

BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,  
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
RA MCMURRAY A., MCINTOSH B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
RA RIFFEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SWALDON N., SMITH A.,  
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,  
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.,  
RL NATURE 368:32-38(1994).  
RA ENBL; AL021481; E1247420; --  
SQ SEQUENCE 344 AA; 33855 MW; FFAE332D6 CRC32;

```

Query Match      75.48; Score 46; DB 5; Length 344;
Best Local Similarity 60.08; Pred. No. 3.26e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db      260 QENPEAPERY 269
QY      1 QENPDSSEPV 10
      |||||

```

|        |  |
|--------|--|
| RESULT | 3  |
| ID     | 002665   |
| AC     | PRELIMINARY; PRT; 590 AA.                                      |
| DT     | 01-JUL-1997 (TREMREL.. 04, CREATED)                            |
| DT     | 01-JUL-1997 (TREMREL.. 04, LAST SEQUENCE UPDATE)               |
| DT     | 01-AUG-1998 (TREMREL.. 07, LAST ANNOTATION UPDATE)             |
| DE     | SODIUM-D-GLUCOSE COTRANSPORTER.                                |
| OS     | ORYZCTOLAGUS CUNICULUS (RABBIT).                               |
| OC     | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA; |
| CC     | EUTHERIA; LAGOMORPHA.  |
| RN     | [1]  |
| RP     | SEQUENCE FROM N.A.   |
| RA     | REINARDT J., GAMBARTAN S., VEYLL M., KOEPESELL H.;             |
| RL     | SUBMITTED (NOV-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.          |
| DR     | ENBL; X82876; E81419; -.                                       |
| SQ     | SEQUENCE 590 AA; 662110 MW; 4686760C CRC32;                    |

```

Query Match      72.1%   Score 44;   DB 6;   Length 590;
Best Local Similarity 75.0%;   Pred. No. 9.31e+00;
Matches 6;   Conservative 2;   Mismatches 0;   Indels 0;   Gaps 0;

Db      43  DNPDPSTEP 50
QY      2  ENPDSSSEP 9
      : |||||

```

|        |  |
|--------|--|
| RESULT | 4  |
| ID     | O67990 PRELIMINARY; PRT; 628 AA.                                       |
| AC     | O67990;  |
| DT     | O1-AUG-1998 (TREMBREL. 07, CREATED)                                    |
| DT     | O1-AUG-1998 (TREMBREL. 07, LAST SEQUENCE UPDATE)                       |
| DT     | O1-AUG-1998 (TREMBREL. 07, LAST ANNOTATION UPDATE)                     |
| DE     | METALLOPROTEASE (FRAGMENT).  |
| VNC.   |  |
| GN     | VIBRIO MIMICUS.  |
| OS     | PRORARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS; |
| OC     | VIBRIONACEAE.  |
| RN     | [1].   |
| KP     | SEQUENCE FROM N.A.   |
| RC     | STRAIN=ATCC33653;  |
| RA     | KONG I.S.;   |
| RL     | SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.                   |
| DR     | EMBL; AF004832; G3142333.  |
| KW     | PROFEASE; METALLOPROTEASE.   |
| FT     | NON_TER<br>I   |
| SQ     | SEQUENCE 628 AA; 71233 MW; 58D32E5O CRC32;                             |

```
Query Match          72.1%; Score 44; DB 2; Length 628;
Best Local Similarity 87.5%; Pred. No. 9.31e+00;
Matches             7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```

Db      568 ENPDSEP 575
      |||||
Qy      2 ENPDSSEP 9

RESULT      5
ID      Q20771      PRELIMINARY;      PRT;      1262 AA.
AC      Q20771;
DT      01-NOV-1996 (TREMBLREL. 01, CREATED)
DT      01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT      01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE      F54D5. 5.
DE      F54D5. 5.
OS      CAENORHABDITIS ELEGANS.
OC      EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN      [1]
RN      SEQUENCE FROM N.A.
RA      COLES L., MATTHEWS L.;
RA      SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN      [2]
RN      SEQUENCE FROM N.A.
RA      WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA      BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA      DEAR S., DU Z., DUBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA      HAWKINS T., HILLIER L., JIER M., JOHNSON L., JONES M., KERSHAW J.,
RA      KRISTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA      MCMURRAY A., MORTMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA      RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SWALDON N., SMITH A.,
RA      SONNHAMMER E., STADEN R., SULLSTON J., THIERRY-MIEG J., THOMAS K.,
RA      VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA      WILKINSON-SPROAT J., WHOLDMAN P.;
RA      WILKINSON-SPROAT J., WHOLDMAN P.;
RL      NATURE 368:32-38(1994).
RL      EMBL; Z66513; G1041333; -.
SQ      SEQUENCE      1262 AA; 144630 MW; 7C2D2904 CRC32;

```

```
Query Match      72.1%   Score 44; DB 5; Length 1262;
Best Local Similarity 30.0%; Pred. No. 9.3le+00;
Matches 3; Conservative 7; Mismatches 0; Indels 0; Caps 0;
```

```

RESULT      6
ID          QI4421 PRELIMINARY; PRT;    105 AA.
AC          QI4421;
DT          01-NOV-1996 (TREMBREL. 01, CREATED)
DT          01-NOV-1996 (TREMBREL. 01, LAST SEQUENCE UPDATE)
DT          01-JUN-1998 (TREMBREL. 06, LAST ANNOTATION UPDATE)
DE          GLYCOPHORIN MZ (FRAGMENT).
GN          GYP A.
OS          HOMO SAPIENS (HUMAN).
OC          EUKARYOTA; METAZOA; CHORDATA; TETRAPODA; MAMMALIA;
OC          EUTHERIA; PRIMATES.
RN          [1]
RP          SEQUENCE FROM N.A.
RX          MEDLINE; 93186803.
RX          HUANG C.H., REID M.E., BLUMENFELD O.O.;
RL          J. BIOL. CHEM. 268:4945-4952(1993).
DR          ENBL; L07253; G183325; -
DR          PROSITE; PS00312; GLYCOPHORIN_A; 1.
KW          ALTERNATIVE SPLICING.
FT          CHAIN        20 >105 POTENTIAL.
FT          NON_TER     105
SQ          SEQUENCE   105 AA; 111086 MW;  2FAB4CCC CRC32;
```

```

Query Match      70.5%; Score 43; DB 4; Length 105;
Best Local Similarity 62.5%; Pred. No. 1.56e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Caps 0;

Db      98 ENPETSDF 105
Qy      2 ENPDSSEP 9
      |||:|:|

```

```

RESULT 7
ID Q99737 PRELIMINARY; PRT; 129 AA.
AC Q99737
DT 01-MAY-1997 (TREMREL. 03, CREATED)
DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE LAMININ ALPHA 4 CHAIN.
GN LAMA4*-1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RA XIAO S., LUX M.L., REEVES R., HUDSON T.J., FLETCHER J.A.;
RL AM. J. PATHOL. 0:0-0(0).
DR EMBL; U77706; G1684837; -.
SQ SEQUENCE 129 AA; 13462 MW; 05562347 CRC32;

Query Match 70.5%; Score 43; DB 4; Length 129;
Best Local Similarity 55.6%; Pred. No. 1.56e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 45 QDPPETSEP 53
QY 1 QENPDSSEP 9

RESULT 8
ID Q20474 PRELIMINARY; PRT; 158 AA.
AC Q20474
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE F46F6.3.
OS CAENORHABDITIS ELGANS.
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RA COTTAGE A.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 368:32-38(1994).
DR EMBL; 250028; G897703; -.
SQ SEQUENCE 158 AA; 18019 MW; C3A480F2 CRC32;

Query Match 70.5%; Score 43; DB 5; Length 158;
Best Local Similarity 55.6%; Pred. No. 1.56e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 83 QHPEDSEP 91
QY 1 QENPDSSEP 9

RESULT 9
ID Q24743 PRELIMINARY; PRT; 184 AA.
AC Q24743
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)

```

```

DE EGP-1 PRECURSOR.
GN EGP-1.
OS DROSOPHILA VIRILIS (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VIRILIS; TISSUE-SALIVARY GLAND;
RA THUROFF E., STOEVEN S., KRESS H.;
RL MECH. DEV. 37:81-93(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VIRILIS; TISSUE-SALIVARY GLAND;
RX MEDLINE; 90276249.
RA SWIDA U., LUCKA L., KRESS H.;
RL DEVELOPMENT 108:269-280(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-VIRILIS; TISSUE-SALIVARY GLAND;
RX MEDLINE; 90384577.
RA KRESS H., SWIDA U.;
RL NATURWISSENSCHAFTEN 77:317-324(1990).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-VIRILIS; TISSUE-SALIVARY GLAND;
RA JARRIN HENTSCHEL A., THUROFF E., TISCHENDORF B., KRESS H.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-VIRILIS; TISSUE-SALIVARY GLAND;
RX MEDLINE; 94200049.
RA KRESS H.;
RL CHROMOSOMA 102:734-742(1993).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN-VIRILIS; TISSUE-SALIVARY GLAND;
RX MEDLINE; 90276248.
RA KRESS H., LUCKA L., SWIDA U., THUROFF E., KLEMM U.;
RL DEVELOPMENT 108:261-267(1990).
DR EMBL; Z49942; G887426; -.
DR FLYBASE; FBgn0005594; Dvir\Egpl.
KW SIGNAL.
FT SIGNAL 1 17 POTENTIAL.
FT SIGNAL 1 18 POTENTIAL.
SQ SEQUENCE 184 AA; 20567 MW; C0E0E0FB CRC32;

Query Match 70.5%; Score 43; DB 5; Length 184;
Best Local Similarity 50.0%; Pred. No. 1.56e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 77 BESPEDEEPV 86
QY 1 QENPDSSEP 10

RESULT 10
ID Q27033 PRELIMINARY; PRT; 480 AA.
AC Q27033
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMREL. 01, LAST ANNOTATION UPDATE)
DE MEMBRANE PROTEIN.
OS THEILERIA PARVA.
OC EUKARYOTA; PROTOZOA; APICOMPLEXA; SPOROZOA; COCCIDIA; PIROPLASMIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MUGUGA;
RX MEDLINE; 94088665.
RA BAYLIS H.A., ALLSOPP B.A., HALL R., CARRINGTON M.;
RL MOL. BIOCHEM. PARASITOL. 61:171-178(1993).
DR EMBL; L06323; G310893; -.
KW MEMBRANE.
SQ SEQUENCE 480 AA; 52388 MW; 1EDC7723 CRC32;

```

Query Match 70.5%; Score 43; DB 5; Length 480;  
 Best Local Similarity 40.0%; Pred. No. 1.56e+01;  
 Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 QOQPDTPQPI 106  
 QY 1 QENPDSSEPV 10

RESULT 11  
 ID P97523 PRELIMINARY; PRT; 1382 AA.  
 AC P97523;  
 DT 01-MAY-1997 (TREMREL. 03, CREATED)  
 DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
 DE HGF RECEPTOR PRECURSOR.  
 GN C-MET.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-LIVER;  
 RA WALLONIUS V., RAHET H., SRTIC S., EXBERG S., HELOU K., QIU Y.,  
 RA LEVAN G., CARLSSON B., ISAKSSON O., NAKAMURA T., JANSSON J.O.;  
 RL MAMM. GENOME 8:661-667(1997).  
 DR EMBL; X96786; E238809;  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PFAM; PF00069; pkinase.  
 KW SIGNAL.  
 FT SIGNAL. 1 24 POTENTIAL.  
 FT CHAIN 25 1382 HGF RECEPTOR.  
 SQ SEQUENCE 1382 AA; 153940 MW; ED5D6941 CRC32;

Query Match 70.5%; Score 43; DB 11; Length 1382;  
 Best Local Similarity 60.0%; Pred. No. 1.56e+01;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 349 QSKPDSAEPM 358  
 QY 1 QENPDSSEPV 10

RESULT 12  
 ID P97579 PRELIMINARY; PRT; 1382 AA.  
 AC P97579;  
 DT 01-MAY-1997 (TREMREL. 03, CREATED)  
 DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)  
 DE HEPATOCYTE GROWTH FACTOR RECEPTOR.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-KIDNEY;  
 RA LIU Y., TOLBERT E.M., SUN A.M., DWORKIN L.D.;  
 RL AM. J. PHYSIOL. 271:0-0(1996).  
 DR EMBL; U65007; G1679660;  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PFAM; PF00069; pkinase.  
 SQ SEQUENCE 1382 AA; 153750 MW; 500939CA CRC32;

Query Match 70.5%; Score 43; DB 11; Length 1382;  
 Best Local Similarity 60.0%; Pred. No. 1.56e+01;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 349 QSKPDSAEPM 358  
 QY 1 QENPDSSEPV 10

RESULT 13  
 ID Q99483 PRELIMINARY; PRT; 81 AA.  
 AC Q99483;  
 DT 01-MAY-1997 (TREMREL. 03, CREATED)  
 DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMREL. 05, LAST ANNOTATION UPDATE)  
 DE GLYCOSYLTRANSFERASE (FRAGMENT).  
 GN ABO.

OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 96437053.  
 RA OGASAWARA K., YABE R., UCHIKAWA M., SAITOU N., BANNAI M., NAKATA K.,  
 RA TAKENAKA M., FUJISAWA K., ISHIKAWA Y., JUJI T., TOKUNAGA K.;  
 RL BLOOD 88:2732-2737(1996).  
 DR EMBL; D82841; G1783220;  
 KW TRANSFERASE.  
 FT NON\_TER 1 81  
 FT NON\_TER 81 81  
 SQ SEQUENCE 81 AA; 8346 MW; 32FC1431 CRC32;

Query Match 68.9%; Score 42; DB 4; Length 81;  
 Best Local Similarity 60.0%; Pred. No. 2.58e+01;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 72 QEPGPGPEPV 81  
 QY 1 QENPDSSEPV 10

RESULT 14  
 ID P91281 PRELIMINARY; PRT; 139 AA.  
 AC P91281;  
 DT 01-MAY-1997 (TREMREL. 03, CREATED)  
 DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
 DE COSMID F27C1.  
 GN F27C1.4.  
 OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE; 94150718;  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
 RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAYTON M.,  
 RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
 RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,  
 RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
 RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
 RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWKNEE R., SMAILDON N., SMITH A.,  
 RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,  
 RA VAUDIN M., VAUGHAN K., WATERSTON R., WAYSON A., WEINSTOCK L.,  
 RA WILKINSON-SPROAT J., WOHLDMAN P.;  
 RL NATURE 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA WU X., LE T.T.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA WATERSTON R.;  
 RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; U80441; G1703603;  
 SQ SEQUENCE 139 AA; 14965 MW; 64437070 CRC32;

Query Match 68.9%; Score 42; DB 5; Length 139;  
 Best Local Similarity 50.0%; Pred. No. 2.58e+01;

Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 111 HENDETGP1 120

Qy 1 QENPDSSEPV 10

RESULT 15

ID Q39754 PRELIMINARY; PRT; 156 AA.

AC Q39754;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE GRP1.

OS FAGUS SYLVATICA.

OC EURARYOTA; PLANTAE; EMBRYOBIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;

OC HAMAMELIDAE; FAGALES; FAGACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-SEEDS:

RA NICOLAS C., NICOLAS G., RODRIGUEZ D.;

RL PLANT MOL. BIOL. 36:487-491(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE-SEEDS:

RA NICOLAS C., NICOLAS G., RODRIGUEZ D.;

RL PLANT CELL PHYSIOL. 38:1303-1310(1997).

DR EMBL; X98339; E249668; -.

SQ SEQUENCE 156 AA; 15164 MW; 4B1907FD CRC32;

Query Match

Best Local Similarity 68.9%; Score 42; DB 10; Length 156;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 31 EKPESNPV 40

Qy 1 QENPDSSEPV 10

Search completed: Thu May 13 15:46:32 1999

Job time : 34 secs.

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 14, 1999, 01:16:20 ; Search time 1151.43 seconds  
(without alignments)  
7587.411 Million cell updates/sec

Title: US-09-040-485-1  
Perfect score: 2442  
Sequence: 1 CGGAGCTTGAAGGACACAA.....GGTTAACTTTAAATATTTT 2442

Scoring table: IDENTITY\_NUC

Searched: 808301 seqs, 1788773984 residues

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_in.\*
- 4: gb\_cm.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl1.\*
- 9: gb\_pl2.\*
- 10: gb\_pr1.\*
- 11: gb\_pr2.\*
- 12: gb\_pr3.\*
- 13: gb\_ro.\*
- 14: gb\_st.\*
- 15: gb\_sy.\*
- 16: gb\_un.\*
- 17: gb\_v1.\*
- 18: gb\_v2.\*
- 19: em\_ba.\*
- 20: em\_fun.\*
- 21: em\_hum1.\*
- 22: em\_hum2.\*
- 23: em\_in.\*
- 24: em\_om.\*
- 25: em\_or.\*
- 26: em\_ov.\*
- 27: em\_pat.\*
- 28: em\_ph.\*
- 29: em\_pl.\*
- 30: em\_ro.\*
- 31: em\_sy.\*
- 32: em\_un.\*
- 33: em\_v1.\*
- 34: em\_v2.\*
- 35: em\_sts.\*
- 36: gb\_ba1.\*
- 37: gb\_ba2.\*
- 38: gb\_pl1.\*
- 39: gb\_pl2.\*
- 40: gb\_pr1.\*
- 41: gb\_pr2.\*
- 42: gb\_pr3.\*
- 43: gb\_sts.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|-------|--------|----|----|-------------|
| -----      |       |       |       |        |    |    |             |

|    |       |      |        |    |           |                    |
|----|-------|------|--------|----|-----------|--------------------|
| 1  | 825.4 | 33.8 | 2449   | 10 | HSU03109  | U03109 Human aspar |
| 2  | 825.4 | 33.8 | 2449   | 40 | HSU03109  | U03109 Human aspar |
| 3  | 823.8 | 33.7 | 2324   | 11 | S83325    | S83325 aspartyl(as |
| 4  | 823.8 | 33.7 | 2324   | 41 | S83325    | S83325 aspartyl(as |
| 5  | 342.6 | 14.0 | 2739   | 4  | BOVASBHY  | M91213 Bos taurus  |
| 6  | 183   | 7.5  | 1777   | 4  | CFU38414  | U38414 Canis famil |
| 7  | 97.4  | 4.0  | 2069   | 27 | E10125    | E10125 DNA encodin |
| 8  | 97.4  | 4.0  | 3399   | 27 | E10126    | E10126 DNA encodin |
| 9  | 95.2  | 3.9  | 1686   | 27 | E08995    | E08995 DNA encodin |
| 10 | 87.6  | 3.6  | 380    | 43 | G23118    | G23118 human sts w |
| 11 | 84.4  | 3.5  | 86     | 6  | AR003317  | AR003317 Sequence  |
| 12 | 79.4  | 3.3  | 112930 | 17 | HSV3PRGEN | X64346 Herpesvirus |
| 13 | 79.4  | 3.3  | 43658  | 17 | HSV3PRGEN | M86409 Herpesvirus |
| 14 | 79.4  | 3.3  | 3720   | 17 | S76368    | S76368 ORF 5' of E |
| 15 | 77.6  | 3.2  | 188324 | 18 | AC005505  | AC005505 *** SEQUE |
| 16 | 71    | 2.9  | 116884 | 18 | PFMALIPI  | AL031744 Plasmodiu |
| 17 | 69.6  | 2.9  | 1324   | 3  | PFRESARI  | X05182 P.falciparu |
| 18 | 69.6  | 2.9  | 4591   | 6  | A00661    | A00661 P.falciparu |
| 19 | 68.6  | 2.8  | 996    | 27 | E08996    | E08996 DNA encodin |
| 20 | 68.6  | 2.8  | 585    | 27 | E08997    | E08997 DNA encodin |
| 21 | 68.2  | 2.8  | 5420   | 3  | AF056936  | AF056936 Plasmodiu |
| 22 | 68    | 2.8  | 7218   | 6  | I66494    | I66494 Sequence 14 |
| 23 | 67.8  | 2.8  | 2384   | 3  | PFAGAR    | J03998 P.falciparu |
| 24 | 67.4  | 2.8  | 3777   | 9  | SPBC215   | AL033534 S.pombe c |
| 25 | 67.4  | 2.8  | 3777   | 39 | SPBC215   | AL033534 S.pombe c |
| 26 | 67    | 2.7  | 8574   | 3  | AF057019  | AF057019 Dictyoste |
| 27 | 66.8  | 2.7  | 914    | 3  | PFRESA    | X55124 P.falciparu |
| 28 | 65.6  | 2.7  | 4050   | 3  | DDPROKING | Z37981 D.discoideu |
| 29 | 65    | 2.7  | 4739   | 3  | PFAGLURPA | M59706 P.falciparu |
| 30 | 65    | 2.7  | 2349   | 6  | A00764    | A00764 Synthetic p |
| 31 | 64.2  | 2.6  | 42141  | 3  | CEY54G9A  | AL032648 Caenorhab |
| 32 | 64.2  | 2.6  | 300197 | 18 | CEY54G9   | Z98869 Caenorhabdi |
| 33 | 64    | 2.6  | 12029  | 3  | AE001373  | AE001373 Plasmodiu |
| 34 | 64    | 2.6  | 334    | 8  | HVC39SAT  | Z50100 H.vulgare G |
| 35 | 64    | 2.6  | 334    | 38 | HVC39SAT  | Z50100 H.vulgare G |
| 36 | 62    | 2.5  | 2941   | 5  | XLU85969  | U85969 Xenopus lae |
| 37 | 62    | 2.5  | 397    | 43 | G37798    | G37798 garp plasm  |
| 38 | 61.6  | 2.5  | 3763   | 5  | PMU19361  | U19361 Petromyzon  |
| 39 | 61    | 2.5  | 3337   | 6  | I23337    | I23337 Sequence 1  |
| 40 | 59.6  | 2.4  | 15421  | 3  | PFCOMPRA  | X95275 P.falciparu |
| 41 | 59.6  | 2.4  | 631    | 3  | PFPPFK51  | X53026 P.falciparu |
| 42 | 59.6  | 2.4  | 2967   | 3  | PFSC03115 | AL021887 Plasmodiu |
| 43 | 59.6  | 2.4  | 3010   | 5  | XLU85970  | U85970 Xenopus lae |
| 44 | 59.6  | 2.4  | 40873  | 18 | CEH04109  | Z92848 Caenorhabdi |
| 45 | 59    | 2.4  | 15148  | 3  | AE001393  | AE001393 Plasmodiu |

ALIGNMENTS

|            |   |          |         |      |     |             |
|------------|---|----------|---------|------|-----|-------------|
| RESULT 1   | HSU03109  | HSU03109 | 2449 bp | mrna | PRI | 30-NOV-1995 |
| LOCUS      | Human aspartyl beta-hydroxylase mRNA, complete cds.   |          |         |      |     |             |
| DEFINITION | U03109  |          |         |      |     |             |
| ACCESSION  | g458031   |          |         |      |     |             |
| NID        |   |          |         |      |     |             |
| KEYWORDS   | human.  |          |         |      |     |             |
| SOURCE     | Homo sapiens  |          |         |      |     |             |
| ORGANISM   | Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.                            |          |         |      |     |             |
| REFERENCE  | 1 (bases 1 to 2249)   |          |         |      |     |             |
| AUTHORS    | Korioti, F., Gieffers, C. and Frey, J.  |          |         |      |     |             |
| TITLE      | Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase   |          |         |      |     |             |
| JOURNAL    | Gene 150 (2), 395-399 (1994)  |          |         |      |     |             |
| MEDLINE    | 95121937  |          |         |      |     |             |
| REFERENCE  | 2 (bases 1 to 2449)   |          |         |      |     |             |
| AUTHORS    | Korioti, F.   |          |         |      |     |             |
| TITLE      | Direct Submission   |          |         |      |     |             |
| JOURNAL    | Submitted (03-NOV-1993) Korioth F., Fakultät fuer Chemie-Biochemie II, Universitaet Bielefeld, Universitaetsstrasse 25, Bielefeld, 33615, Germany |          |         |      |     |             |



| 3' UTR                | 778 a  | 507 c               | 655 g     | 509 t        |  |
|-----------------------|--|---------------------|-----------|--------------|--|
| BASE COUNT            | 2352   | .2449               |           |              |  |
| ORIGIN                |  |                     |           |              |  |
| Query Match           | 33.8%  | Score 825.4;        | DB 10;    | Length 2449; |  |
| Best Local Similarity | 99.3%  | Pred. No. 1.5e-132; |           |              |  |
| Matches 829;          | Conservative 0;  | Mismatches 6;       | Indels 0; | Gaps 0;      |  |
| QY 5                  | AGCTTGAAGGACACAAGATGGGAGGAAGCGGACTCTCAGGAACCTCATCTTCACGT     | 64                  |           |              |  |
| DB 187                | AGCATGGAGGACACAAGATGGGAGGAAGCGGACTCTCAGGAACCTCATCTTCACGT     | 246                 |           |              |  |
| QY 65                 | GGTTTATGGTGATGTCATCTCGGCGCTCTGGACATCTAGCTGTCTGGTTGGTTGATC    | 124                 |           |              |  |
| DB 247                | GGTTTATGGTGATGTCATCTCGGCGCTCTGGACATCTAGCTGTCTGGTTGGTTGATC    | 306                 |           |              |  |
| QY 125                | TTCATTGACATGAGGAAGTTCTTAGGAAGAACTAGGAATCTATGATGCTGGTGGATGGAG | 184                 |           |              |  |
| DB 307                | TTCATTGACATGAGGAAGTTCTTAGGAAGAACTAGGAATCTATGATGCTGGTGGTGGAG  | 366                 |           |              |  |
| QY 185                | ATTTTTCATGTGGATGATGCCAAAGTTTATTAGGACTTAAAGAGAGATCTACTTCAGAC  | 244                 |           |              |  |
| DB 367                | ATTTTTCATGTGGATGATGCCAAAGTTTATTAGGACTTAAAGAGAGATCTACTTCAGAC  | 426                 |           |              |  |
| QY 245                | CAGCAGTCCCGCCAGGAAGGCTGAGCCACACACTGAGCCGAGGAGCAGGTCCTCTGG    | 304                 |           |              |  |
| DB 427                | CAGCAGTCCCGCCAGGAAGGCTGAGCCACACACTGAGCCGAGGAGCAGGTCCTCTGG    | 486                 |           |              |  |
| QY 305                | AGCGAAGACCCCAAGATATCGAAGATGAAGCAAGAAACAAATTCAGTCCCTTCTCCATG  | 364                 |           |              |  |
| DB 487                | AGCGAAGACCCCAAGATATCGAAGATGAAGCAAGAAACAAATTCAGTCCCTTCTCCATG  | 546                 |           |              |  |
| QY 365                | AAATGGTACACGACGAACATGTTGAGGAGAGACTTTGCAACAGAGAGATGGACCCACAG  | 424                 |           |              |  |
| DB 547                | AAATGGTACACGACGAACATGTTGAGGAGAGACTTTGCAACAGAGAGATGGACCCACAG  | 606                 |           |              |  |
| QY 425                | GAGAACCACAAAGAGGATGATGATTTCTTAGGGGACTGATGATAGATGATGATTTG     | 484                 |           |              |  |
| DB 607                | GAGAACCACAAAGAGGATGATGATTTCTTAGGGGACTGATGATGATGATGATTTG      | 666                 |           |              |  |
| QY 485                | AGACCTTGAACCTGAAGTATCTCATGAAGAAACCGAGCATAGTTACCACTGGAAGAGA   | 544                 |           |              |  |
| DB 667                | AGACCTTGAACCTGAAGTATCTCATGAAGAAACCGAGCATAGTTACCACTGGAAGAGA   | 726                 |           |              |  |

|                       |   |
|-----------------------|---|
| NID                   | g1911651  |
| KEYWORDS              | human hepatoblastoma cell line HepG2.   |
| SOURCE                | Homo sapiens  |
| ORGANISM              | Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 2324)  |
| REFERENCE             | Lavalssiere,L., Jia,S., Nishiyama,M., de la Monte,S., Stern,A.M., Wands,J.R. and Friedman,P.A.  |
| AUTHORS               | Overexpression of human aspartyl(beta-hydroxylase)in hepatocellular carcinoma and cholangiocarcinoma J. Clin. Invest. 98 (6), 1313-1323 (1996)  |
| TITLE                 | GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 179962] from the original journal article. This sequence comes from Fig. 4.   |
| JOURNAL               | Authors note differences between this sequence and that of F. Koriotoh, C. Gieffers, and J. Frey: Gene 150 (2), 393-399 (1994), GenBank U03109.   |
| MEDLINE               | Location/Qualifiers   |
| REMARK                | 1..2324   |
|                       | /organism="Homo sapiens"  |
|                       | /db_xref="taxon:9606"   |
|                       | 12..2288  |
|                       | /gene="aspartyl(beta-hydroxylase), HAAH"  |
|                       | 12..2288  |
|                       | /note="This sequence comes from Fig. 4; HAAH"   |
|                       | /codon_start=1  |
|                       | /product="aspartyl(asparaginyl)beta-hydroxylase"  |
|                       | /db_xref="PID:g1911652"   |
|                       | /translation="MAQRNKAAGSSSSSGSGSTAGSSSPGARRETKHGGKNG RKGLSGTSFTWMTALVGVSVAVVDFVLVDVEYVGLGIIVDADGDGFDDVD AKVLLGKERSTSEPAVPPEAEAPHTEPEQVPEARPONIDEAKEQIOSLHEMVA AEHVGEDLQOEDPTGEPODEDELMDATDVDRFETLEPEVSHEIETSYRVEEVT SQCNQDMEEEMSEQNPDSPEVDEDLRHDTDDTYIQVIEQAAYVEPLENEGIEI TETAPDENPDNSQVIEEVSIFFVEQQEYPPENRTKDDPEQAKVKKKPKILL KDFTKIKAELDAEKLRGKTEEAUNAFLKLVKPQSPRAYRYKAQCEDDOLAEKR RSNEVLGAJETQEVASLPDPADLLKLSLKRRSDROQFLGHMRGSLTLORLVOLF PNDTSLKNDLGCVLLIGDNDNAKVYEVLVSTPNDGFAKYHYGTILKAQNKIAESI PYLKEGESGDPDDBRGTFYHLGDAMVRGNKEAYKWELGHGRHFASFWORSLTIN VNLGAQPWTTPKETGYTELTVLSLERNMKLIRDEGLAMDKAKGLFLPEDENREKGD WSCLTMOQGRNRNACKGAPKTCTLLEKFPETTGCRGOIKYSIHHPGHVPHWPHTG TNCRLRMHLGLVIPKEGCKIRCANEITRWEEGKVLIFDDSFEEHVQDASSFRLIIV DVVHPELTPQORBSLPAI" |
| BASE COUNT            | .753 a .463 c .625 g .483 t   |
| ORIGIN                |   |
| Query Match           | 33.7%; Score 823.8; DB 11; Length 2324;   |
| Best Local Similarity | 99.2%; Pred. No. 2.8e-132;  |
| Matches               | 826; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  |
| Qy                    | 5 AGCTTTGAGGACACAAGAATGGGAGAAAGCGGACTCTCAGGAACCTTCATTCTTCAGGT 64<br>  |
| Db                    | 121 AGCATGGAGGACACAAGAATGGGAGAAAGCGGACTCTCGGAACTTCATTCTTCAGGT 180<br>   |
| Qy                    | 65 GGTTTATGGTGATTTGCATTTGCTGGCGCTCTGGACATCTGTAGCTGTGCTTTGGTTTGA TC 124<br>  |
| Db                    | 181 GGTTTATGGTGATTTGCATTTGCTGGCGCTCTGGACATCTGTAGCTGTGCTTTGGTTTGA TC 240<br>   |
| Qy                    | 125 TTGTTGACTATGAGGAAGTCTTAGSAAACTAGSAACTATGATGCTGATGCTATGATGAG 184<br>   |
| Db                    | 241 TTGTTGACTATGAGGAAGTCTTAGSAAACTAGSAACTATGATGCTGATGCTGATGAG 300<br>   |
| Qy                    | 185 ATTTTGTATGCTGGATGCCAAGTTTTATTAGGACTTTAAAGAGAGATCTACTTCAGAGC 244<br>   |
| Db                    | 301 ATTTTGTATGCTGGATGCCAAGTTTTATTAGGACTTTAAAGAGAGATCTACTTCAGAGC 360<br>   |
| Qy                    | 245 CAGCAGTCCGCCAGAGAGAGGCTGAGCACACACTGAGCCCCGAGAGAGAGGTTTCTCTGTGG 304<br>  |
| Db                    | 361 CAGCAGTCCGCCAGAGAGAGGCTGAGCACACACTGAGCCCCGAGAGAGGTTTCTCTGTGG 420<br>  |

```
QY 305 AGGCGAAGCCAGAAATATCGAGATGAAGCAAAAGAAACAAATTCAGTCCCTCTCCATG 364
Db 421 AGGCGAAGCCAGAAATATCGAGATGAAGCAAAAGAAACAAATTCAGTCCCTCTCCATG 480
QY 365 AAATGGTACACGAGAAATATGTTGAGGGAGAGACATTCGAACAGAGATGGACCCACAG 424
Db 481 AAATGGTACACGAGAAATATGTTGAGGGAGAGACATTCGAACAGAGATGGACCCACAG 540
QY 425 GAGAACCAACAAAGAGGATGATGAGTCTTCTATGCGGACTGATGATGATGATGATGATG 484
Db 541 GAGAACCAACAAAGAGGATGATGAGTCTTCTATGCGGACTGATGATGATGATGATGATG 600
QY 485 AGACCCCTGGAACCTGAAGTATCTCATGAAGAAACCCAGCATGATTCACCGTGGAAAGAGA 544
Db 601 AGACCCCTGGAACCTGAAGTATCTCATGAAGAAACCCAGCATGATTCACCGTGGAAAGAGA 660
QY 545 CAGTTTCAACAGACTGTAATCAGGATATGGAAGAGATGATGTCGACGAGAAATCCAG 604
Db 661 CAGTTTCAACAGACTGTAATCAGGATATGGAAGAGATGATGTCGACGAGAAATCCAG 720
QY 605 ATTCCAGTGAACCTAGTAGAGATGAAGATGTCACCATGATACAGATGATGATGATGATG 664
Db 721 ATTCCAGTGAACCTAGTAGAGATGAAGATGTCACCATGATACAGATGATGATGATGATG 780
QY 665 ACCAAGTCTATGAGGAACAGAGATATGAACCTCTAGAAATGAAGGATAGAAATCA 724
Db 781 ACCAAGTCTATGAGGAACAGAGATATGAACCTCTAGAAATGAAGGATAGAAATCA 840
QY 725 CAGAAGTAACTGCTCCCTGAGGATAATCCTGTAGAAGATTCACAGGTAATGTAGAAG 784
Db 841 CAGAAGTAACTGCTCCCTGAGGATAATCCTGTAGAAGATTCACAGGTAATGTAGAAG 900
QY 785 AAGTAAGCAATTTTCTGTGGAGACAGCAGGAAGTACCACAGATCTTAAG 839
Db 901 AAGTAAGCAATTTTCTGTGGAGACAGCAGGAAGTACCACAGATCTTAAG 955

RESULT 4
S83325 LOCUS S83325 2324 bp mRNA PRI 27-MAR-1997
DEFINITION aspartyl(asparaginyl)beta-hydroxylase [human, hepatoblastoma cell
line HepG2, mRNA Partial, 2324 nt].
ACCESSION S83325
NID 91911651
KEYWORDS human hepatoblastoma cell line HepG2.
SOURCE Homo sapiens
ORGANISM
REFERENCE Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS 1 (bases 1 to 2324)
Lavalsiere, L., Jia, S., Nishiyama, M., de la Monte, S., Stern, A.M.,
Wands, J.R. and Friedman, P.A.
TITLE Overexpression of human aspartyl(asparaginyl)beta-hydroxylase in
hepatocellular carcinoma and cholangiocarcinoma
J. Clin. Invest. 98 (6), 1313-1323 (1996)
JOURNAL 96420598
MEDLINE
REMARK Genbank staff at the National Library of Medicine created this
entry [NCBI gibbs 17962] from the original journal article.
This sequence comes from Fig. 4.
COMMENT Authors note differences between this sequence and that of F.
Korioth, C. Gleffers, and J. Frey: Gene 150 (2), 395-399 (1994),
GenBank U03109.
FEATURES Location/Qualifiers
source 1..2324
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene 12..2288
CDS 12..2288
/gene="aspartyl(asparaginyl)beta-hydroxylase, HAAH"
/note="This sequence comes from Fig. 4; HAAH"
/codon_start=1
/product="aspartyl(asparaginyl)beta-hydroxylase"
```

```
/db_xref="PID:g1911652"
/translation="MAQRNKAQSSSSSGSGSTAGSSSPGARRETTHGKHNG
RGGSLSGSTFTFWVIALGLVWTSVAVVWFDLVVEEVLKGLIYDADGDFDQDD
AKVLGKRSSTSEPAVPEEAPHEPEEVPVPEAEQNIIEDEAEQIQSLDHEWVH
ASHVEDGLQEDGPTGEQDDDEFDMATDVRDFTLEPEVSHETVHEVETV
SDCQDMEMNSEQENPDSPVDEDLHDHDTDDVYQVVEEQAVPEPLENEGLEI
TWTAPDNVEDQSVIIEVEISIPVEEQVPPETNRKTDPEQAKKKPKLL
NFKDTIKALDAAEKLRKGLIEAVNAFKELVRKYPQSPRARGKACEDDLAEKR
RNEVLGAIETQEVASLPDADLLKLSLRSDRQFLGMRGSLTLQRLVOLF
PNDTSLKNDLGVGILLIGDNDNAKVIEEVLSTPNDGFAKHVYGLKAKNQKIAESI
PYLKEGIESDPTDGRFYFLGDAMQVKNKAYKWTGLGKHGFILKASVQWRSYN
VNLGAQPMWTKETGYTELKSLERNWKLIRDEGLAVMDKAGLFLPDELENREKGD
WSOFTLWOQRNENAKGAPKPTLTLEFPEPTTTCRRGQIKVSIHMPGTHVMPHTGP
TNCRLMHGLVLPKEGCKIRCANETRTWEEKVLIFDDSFHEVWQDASSFLIIV
DWHBELTPQQRRLPAL"
BASE COUNT 753 a 463 c 625 g 483 t
ORIGIN

Query Match 33.7%; Score 823.8; DB 41; Length 2324;
Best Local Similarity 99.2%; Pred. No. 2.8e-132;
Matches 828; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 5 AGCTTGAAGGACACAAAGATGGGAGGAGCGGACTCTCAGGAACCTTCATTCTTCACGT 64
Db 121 AGCATGGAGGACACAAAGATGGGAGGAGCGGACTCTCGGAACCTTCATTCTTCACGT 180
QY 65 GGTATTGTTGATTGCTGCGGCTCTGGACATCTGTAGCTGTCTGTTGTTGTTGATC 124
Db 181 GGTATTGTTGATTGCTGCGGCTCTGGACATCTGTAGCTGTCTGTTGTTGTTGATC 240
QY 125 TTGTTGACTATGAGGAAGTCTAGGAAACTAGGAATCTATGATGCTGTGATGATGAGG 184
Db 241 TTGTTGACTATGAGGAAGTCTAGGAAACTAGGAATCTATGATGCTGTGATGATGAGG 300
QY 185 ATTTTGTATGTTGATGATGCCAAAGTTTATTAGGACTTTAAAGAGAGATCTACTTTCAGAGC 244
Db 301 ATTTTGTATGTTGATGATGCCAAAGTTTATTAGGACTTTAAAGAGAGATCTACTTTCAGAGC 360
QY 245 CACAGTCCCGCCAGAGAGGCTGAGCCACACACTGAGCCCGAGGAGCAGGTTCTCTGTGG 304
Db 361 CACAGTCCCGCCAGAGAGGCTGAGCCACACACTGAGCCCGAGGAGCAGGTTCTCTGTGG 420
QY 305 AGGCAGAACCCCAAGATATCGAAGATGAAGCAAAAGAAACAAATTCAGTCCCTCTCCATG 364
Db 421 AGGCAGAACCCCAAGATATCGAAGATGAAGCAAAAGAAACAAATTCAGTCCCTCTCCATG 480
QY 365 AAATGTACACGAGAACATGTTGAGGGAGAGACACTGCAACAGAGATGGACCCACAG 424
Db 481 AAATGTACACGAGAACATGTTGAGGGAGAGACACTGCAACAGAGATGGACCCACAG 540
QY 425 GAGAACCAACAAAGAGGATGATGAGTCTTCTATGCGGACTGATGATGATGATGATGATG 484
Db 541 GAGAACCAACAAAGAGGATGATGAGTCTTCTATGCGGACTGATGATGATGATGATGATG 600
QY 485 AGACCTGGAACCTGAAGTATCTCATGAAGAAACCCAGCATGATTCACCGTGGAAAGAGA 544
Db 601 AGACCTGGAACCTGAAGTATCTCATGAAGAAACCCAGCATGATTCACCGTGGAAAGAGA 660
QY 545 CAGTTTCAACAGACTGTAATCAGGATATGGAAGAGATGATGTCGACGAGAAATCCAG 604
Db 661 CAGTTTCAACAGACTGTAATCAGGATATGGAAGAGATGATGTCGACGAGAAATCCAG 720
QY 605 ATTCCAGTGAACCTAGTAGAGATGAAGATGTCACCATGATACAGATGATGATGATGATG 664
Db 721 ATTCCAGTGAACCTAGTAGAGATGAAGATGTCACCATGATACAGATGATGATGATGATG 780
QY 665 ACCAAGTCTATGAGGAACAGAGATATGAACCTCTAGAAATGAAGGATAGAAATCA 724
Db 781 ACCAAGTCTATGAGGAACAGAGATATGAACCTCTAGAAATGAAGGATAGAAATCA 840
QY 725 CAGAAGTAACTGCTCCCTGAGGATAATCCTGTAGAAGATTCACAGGTAATGTAGAAG 784
Db 841 CAGAAGTAACTGCTCCCTGAGGATAATCCTGTAGAAGATTCACAGGTAATGTAGAAG 900
```

```
Db 841 CAGAACTAACTGCTCCCTCCTGAGATAATCCTGTAGAAAGATTCACAGGTAAATGTAGAAG 900
QY 785 AAGTAAGCAATTTTCTCTGGAAGACAGCAGGAAGTACCACAGATACCTAAAG 839
Db 901 AAGTAAGCAATTTTCTCTGGAAGACAGCAGGAAGTACCACAGAAACAAATAG 955

RESULT 5
LOCUS BOVASBH 2739 bp mRNA MAM 08-DEC-1993
DEFINITION Bos taurus aspartyl (asparaginyl) beta hydroxylase mRNA, complete
cfs.
ACCESSION M91213
NID g162693
KEYWORDS aspartyl (asparaginyl) beta hydroxylase.
SOURCE Bos taurus cDNA to mRNA.
ORGANISM Bos taurus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 2739)
Jia,S., VanDusen,W.J., Diehl,R.E., Kohl,N.E., Dixon,R.A.,
Elliston,K.O., Steirn,A.M. and Friedman,P.A.
cDNA cloning and expression of bovine aspartyl (asparaginyl)
beta-hydroxylase
J. Biol. Chem. 267 (20), 14322-14327 (1992)
JOURNAL 9232546
MEDLINE 2 (bases 1 to 2739)
REFERENCE Friedman,P.A.
AUTHORS Direct Submission
JOURNAL Submitted (15-APR-1992) P.A. Friedman, Merck Sharp and Dohme
Research Laboratories, West Point, PA 19486 USA
FEATURES
source
Location/Qualifiers
1..2739
/organism="Bos taurus"
/db_xref="taxon:9913"
/tissue.type="liver and brain"
189..2453
/codon_start=1
/product="aspartyl (asparaginyl) beta hydroxylase"
/db_xref="PID:g162694"
/translation="MAPRKNAGGGSSSSSGSPGTCGTGGSSSPGARRETQKGL
KNGKGGSGSGSFFTFWVYIALLGWTSVAVWFELVDYEEVLAKDFRYNLSEVLQ
KGLGIYDAGDGDDEVDKVLGLGKERPAFTVPPEADMPLEDOVLESPPGQN
IEDEVEYQVSLDEIVISPEGLPQEPGAELQPDHDFVGSDDADRYEPMGTGA
VHEETEDYHIEETASPAYSDMEDMYEQENPDSPEPVVDADERTYQETDVTYRD
YDEQHDVNSNTILEEHPMPAEQOEVPETNKKADPEKGGKVKKKPKLKNKFD
TKIAELDAAEKLRKGTETEEAVNAFEELVRKYPOSPGARYKQACEDDLAEKRSNE
ILRAIETQEASLDPADTDVLKLSLRSDRQFLGHRGSLTLQLVLPDDT
ALKNDLGVYLIGDNDGAKVYEEVLPNDGFAKHVYFILKAQNKIASSIPLYK
EGIESDPGTDGTFYHLGDAMQVRGNKEAYRWELHQRGHFASVWQSLYNVQGL
KAQPMWPKETITELVSLERNWLIRDEGLAAMDRTGLFLPEDENLRKGDWSOF
TLWQGRKNENACKAPKTSLLDKFPETTCGRGQIKYSIMHPGTHVPHGTNCR
LRMHGLVLPKEGCKIRCANETRTWEEGKVLIFDSDFEHVEWQDAASFLIFIVDWH
PELTPHQRSLPAI"
BASE COUNT 801 a 636 c 749 g 553 t
ORIGIN

Query Match 14.0%; Score 342.6; DB 4; Length 2739;
Best Local Similarity 67.4%; Pred. No. 6.8e-50;
Matches 607; Conservative 0; Mismatches 224; Indels 70; Gaps 6;

QY 3 GGAGCTTGAGGACACAGAATGGGAGGAAGCGGACTCTCAGGAACCTCAATCTTCAC 62
Db 305 GAACAGGAGGACTCAAGATGGGAGGAGGAGGACITTCCTGGAGTTCATTTTCAC 364
QY 63 GTGGTTTATGGTATTCATTCGTGGCGGTCTGGACATCTGTAGCTGTCTGTTTGA 122
Db 365 ATGGTTTATGGTATTCATTCGTGGCGGTCTGGACATCTGTAGCTGTCTGTTTGA 424
QY 123 TCTGTGTACTATGAGGAG----- 143
```

```
Db 425 TCTTGTGATTATCAAGAAGTGTAGCCAAAGCAAGGACTTCGCTTATAACTTGTGAGA 484
QY 143 -----TTCTAGGAAAACCTAGGAATCTATGATGCTGATGGTATGGAGATTTTGTGGA 197
Db 485 GGTACTTCAAGGAAACCTTGGAAATCTATGATGCTGATGGTATGGAGATTTTGTGGA 544
QY 198 TGATGCCAAAAGTTTATTAGGACTTTAAAGAGAGATCTACTTCAGAGCCAGGACTCCCGCC 257
Db 545 TGATGCCAAAAGTTTATTAGGCTTTAAAGAAAACCTGCTCCAAAGCAACAGTGCAC 604
QY 258 AGAAGAGGCTGAGCCACACACTGAGCCCGAGGAGCAGGTTCCTCTGGAGGAGCAACCCCA 317
Db 605 AGAAGAGGCGAGACATGTACCTTGGCTGGAAGATCAGTCTCTGAGAGCCCGAGGAGCTCA 664
QY 318 GAATATCGAAGATGAAGCAAAAGAAACAAATTCAGTCCCTTCTCCATGAAATGTACAGC 377
Db 665 GAATATTGAAGATGAAGTATATATGAACAAGTTTCACTCACTT---GATGAAACAGTATACT 721
QY 378 AGACATGTTGGGAGCAAGACTTGCACAGAGATGTGGCCCAAGAGGAGGAGGAGGAGGAGGAG 437
Db 722 AGAACCT-----GGAGAGAACCTGCCCCAAGAACCAAGAGGAGGAGGAGGAGGAGGAG 775
QY 438 AGAGGATGATGAGTTCCTTATGGCGACTGATGATGATGATGATGATGATGATGATGATGATG 497
Db 776 GCCCGAGGACCATCTCTTTGTAGGAAGTGTACCATATAGACAGATACCAACCATTTGGGAC 835
QY 498 TGAAGTATCTATGAAGAAACCCGAGCATAGTTACACGCTGGAAGAGACAGAGTTCACAGA 557
Db 836 TGGAGCCGTTTCATGAAGAAACTGAAGATAGTACCATATAGACAGACAGCTTCGCCAGC 895
QY 558 CTGTAATCAGATATGAAGAGATGTCTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 617
Db 896 CTACAGTCAAGATATGAAGACATGATGTATGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 955
QY 618 AGTAGTAGAAGA-----TGAAGATTGCACCATGATACAGATGATGATGATGATGATGATGATG 671
Db 956 AGTAGTAGTAGATGATGCTGAAGAACATACCAAGAACAGATGATGATGATGATGATGATGATG 1015
QY 672 CTATGAGGACAAACAGCAGTATATGAACCTTAGAAAATGAAGGATAGAAAATCACAGAGT 731
Db 1016 CTATGATGAACAAG-----ATCATGCTGTAGATAATTCACACACATTTTAGAGAGAC 1069
QY 732 AACTGCTCCCTCGAGGATTAATCTCTAGAGATTCACAGTAAATTTAGAGAGAGTAAAG 791
Db 1070 GCATATGCCCTCGAGAGAACAGCAAGAGTACCACAGAAA---CAAATAGAGAG 1125
QY 792 CATTTTCTCTGTGGAAGAACAGCAGGAGTACCCACAGATACITTAAGCTTCAAAAGAC 851
Db 1126 CAGATGAGCCAGGAAAAAAGGAAAGTGAAGAAAAAGAGCCCTAAATATTAATAAT 1185
QY 852 T 852
Db 1186 T 1186

RESULT 6
CFU38414
LOCUS Canis familiaris junctional sarcoplasmic reticulum protein mRNA,
DEFINITION complete cds.
ACCESSION U38414
NID g1163912
KEYWORDS dog.
SOURCE Canis familiaris
ORGANISM Canis familiaris
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 1777)
Jones,L.R., Zhang,L., Sanborn,K., Jorgensen,A.O. and Kelley,J.
Purification, primary structure, and immunological characterization
of the 26-kDa calsequestrin binding protein (junctin) from cardiac
junctional sarcoplasmic reticulum
J. Biol. Chem. 270 (51), 30787-30796 (1995)
```

MEDLINE 96107245  
 REFERENCE 2 (bases 1 to 1777)  
 AUTHORS Jones,L.R., Zhang,L., Sanborn,K., Jorgensen,A.O. and Kelley,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-OCT-1995) Larry R. Jones, Medicine, Krannert Inst.  
 Cardiology, 1111 W. 10 Street, Indianapolis, IN 46202, USA  
 FEATURES Location/Qualifiers  
 source 1..1777  
 /organism="Canis familiaris"  
 /db\_xref="taxon:9615"  
 /tissue\_type="heart"  
 89..721  
 /note="junction"  
 /codon\_start=1  
 /function="calcineurin binding protein"  
 /product="junctional sarcoplasmic reticulum protein"  
 /db\_xref="pid:g1163913"  
 /translation="MAPEIKHGKNGKGLSGSSFTFMVIALIGVYTSVAVWF  
 DLVDYEVGLKGVADGDVDDDKVLLGPGGVAKRKTAKVKELTKELKKE  
 KETESRKNENRKKRKKEDERDKKIADADISRKESPKGKDKREKNVGLDKS  
 AKAKESRKRSTNVKDASSASTASRDKDDTKEGKTSKRTHSAKGNQKRKN"  
 polyA\_site 1777  
 /note="44 A nucleotides"  
 BASE COUNT 552 a 331 c 363 g 531 t  
 ORIGIN

Query Match 7.5%; Score 183; DB 4; Length 1777;  
 Best Local Similarity 90.7%; Pred. No. 1.6e-22;  
 Matches 195; Conservative 0; Mismatches 20; Indels 0; Gaps 0;  
 QY 5 AGCTTGAGGACACAAATGGGAGGAGGCGACCTCAGGAACCTTCATCTTCACGT 64  
 DB 105 AGCATGAGGACACAAATGGGAGGAGGAGGACCTTCGGAAGTTCATTTTCACAT 164  
 QY 65 GGTATTGATGCTATGCTGGCGCTCTGGACATCTGTAGCTGTCTTTGGTTGATC 124  
 DB 165 GGTATTGATGCTATGCTGGCGCTCTGGACATCTGTAGCTGTCTTTGGTTGATC 224  
 QY 125 TTGTTGACTATGAGGAAGTCTTAGGAAACTAGGAATCTATGATGCTGATGGAG 184  
 DB 225 TTGTTGATTAGGAAGTCTTAGGAAACTAGGAGTCTATGATGCTGATGGAG 284  
 QY 185 ATTTGATGCTGATGATGCCAAAGTTTATTAGGA 219  
 DB 285 ATTTGATGCTGATGATGCCAAAGTTTATTAGGA 319

RESULT 7  
 ID E10125 standard; RNA; UNC; 2069 BP.  
 XX AC  
 XX AC  
 XX NI  
 XX d1108462  
 DT 08-OCT-1997 (Rel. 52, Created)  
 DT 08-OCT-1997 (Rel. 52, Last updated, Version 1)  
 XX  
 DE DNA encoding an immunogenicity protein.  
 XX  
 KW JP 1995284392-A/1.  
 XX  
 OS unidentified  
 OC unclassified.  
 XX  
 RN [1]  
 RP 1-2069  
 RA Doi H., Nagakuchi Y., Tanaka Y., Fujisaki Y.;  
 RT "GENE CLONE MANIFESTING CHICKEN LEUCOCYTOZOON IMMUNOGENIC PROTEIN AND  
 RT GENE RECOMBINANT VACCINE AGAINST CHICKEN LEUCOCYTOZOON";  
 RL Patent number JP 1995284392-A/1. 31-OCT-1995.  
 RL DOUBUTSUYOU SEIBUTSUGAKUTEKI SEIZAI KYOKAI, KITASATO INST:THE.

XX OS Leucocytozon caulleryi  
 CC PN JP 1995284392-A/1  
 CC PD 31-OCT-1995  
 CC PF 19-APR-1994 JP 1994080643  
 CC PI DOI HIROHITO, NAGAKUCHI YOSHIO, TANAKA YOSHIO, FUJISAKI YUJIRO  
 CC PC C12N15/09,A61K39/015,C12P21/02;  
 CC CC strandedness: Double;  
 CC CC topology: Linear;  
 CC CC key Location/Qualifiers  
 CC FH source 1..2069  
 CC FT /organism="Leucocytozon caulleryi"  
 CC FT /strain="shizuoka"  
 CC FT 1..25  
 CC FT misc\_feature /note="EcoRI adaptor"  
 CC FT misc\_feature 2045..2069  
 CC FT /note="EcoRI adaptor"  
 CC FT misc\_feature 25..2046  
 CC FT /product="immunogenicity protein"  
 XX Key Location/Qualifiers  
 FH source 1..2069  
 FT /organism="unidentified"  
 FT  
 XX Sequence 2069 BP; 1216 A; 167 C; 443 G; 243 T; 0 other;

Query Match 4.0%; Score 97.4; DB 27; Length 2069;  
 Best Local Similarity 46.3%; Pred. No. 7e-08;  
 Matches 320; Conservative 0; Mismatches 371; Indels 0; Gaps 0;  
 QY 147 AGGAAACTAGGAATCTATGATGCTGTGATGGTGGAGATTTTGTGATGGATGCCAA 206  
 DB 1053 AGAAAAGAAAGTACACATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1112  
 QY 207 AGTTTATTAGGACTTAAAGAGAGATCTACTTCAGAGCCAGCAGTCCGCCAGAGAGGC 266  
 DB 1113 AGAAAAGTAAATACATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1172  
 QY 267 TGACCCACACTCAGCCCGAGGAGCAGGTTCTTGTGGAGCCAGACCCCAAGATATCGA 326  
 DB 1173 AGTAACACATGAAG 1232  
 QY 327 AGATGAAG 386  
 DB 1233 ACATGAAG 1292  
 QY 387 TGAGGAGAGAGACTTGCAACAAGAGAGATGGACCCAGGAGAGAGAGAGAGAGAGAGAG 446  
 DB 1293 AGAAG 1352  
 QY 447 TGAGTTTCTTATGCGGACTGATGTAGATGTAGATTTGAGACCTGGAACTGAAGTATC 506  
 DB 1353 AGAAG 1412  
 QY 507 TCATGAAG 566  
 DB 1413 ACATGAAG 1472  
 QY 567 GGATATGGAAGAGAGATGTCTGAGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 626  
 DB 1473 ACATGAAG 1532  
 QY 627 AGATGAAG 686  
 DB 1533 AGAAG 1592  
 QY 687 AGTATATGAACCTCTAGAAATGAAGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 746  
 DB 1593 AGAAG 1652

QY 747 GGATAATCCTGTAGAGATTACACAGGTAAATGTAGAGAGTAAGCATTTTCTCTGTGGA 806  
 ||||| | ||||| | | | | | ||||| | | |  
 Db 1653 AGTAGGAG 1712  
 ||||| | ||||| | | | | | ||||| | | |  
 QY 807 AGAACAGCAGAGAGTACCACAGATCTTAA 837  
 ||||| | ||||| | | | | | ||||| | | |  
 Db 1713 AGAAGAAG 1743

RESULT 8  
 E10126 standard; DNA; UNC; 3399 BP.  
 XX E10126;  
 AC  
 XX  
 NI d1108463  
 XX  
 XX  
 DT 08-OCT-1997 (Rel. 52, Created)  
 DT 08-OCT-1997 (Rel. 52, Last updated, Version 1)  
 XX  
 DE DNA encoding an immunogenicity protein of Leucocytozon caulleryi  
 DE fused to maltose-binding protein.  
 XX

KW JP 1995284392-A/2.

XX unidentified  
 XX unclassified.

RN [1]  
 RP 1-3399  
 RA Doi H., Nagakuchi Y., Tanaka Y., Fujisaki Y.;  
 RT "GENE CLONE MANIFESTING CHICKEN LEUCOCYTOZON IMMUNOGENIC PROTEIN AND  
 RT GENE RECOMBINANT VACCINE AGAINST CHICKEN LEUCOCYTOZON";  
 RL Patent number JP 1995284392-A/2, 31-OCT-1995.  
 RL DOUBTSYOU SEIBUTSUGAKUTEKI SEIZAI KYOKAI, KITASATO INST:THE.

XX OS None  
 CC Artificial sequences.  
 CC PN JP 1995284392-A/2  
 CC PD 31-OCT-1995  
 CC PF 19-APR-1994 JP 1994080643  
 CC PI DOI HIROHITO, NAGAKUCHI YOSHIO, TANAKA YOSHIO, FUJISAKI YUJIRO  
 CC PC C12N15/09 A61K39/015, C12P21/02;  
 CC CC strandedness: Double;  
 CC CC topology: Linear;

CC FH Key Location/Qualifiers  
 CC FT source 1. .3399 /organism="Artificial sequences"  
 CC FT CDS 1. .3399 /product="fusion protein of maltose-binding  
 CC FT protein and an immunogenicity protein"  
 CC FT misc\_feature 1. .1149 /note="maltose-binding protein"  
 CC FT 1150. .1174 /note="EcoRI adaptor"  
 CC FT 1174. .3195 /note="immunogenicity protein"  
 CC FT 3194. .3218 /note="EcoRI adaptor"  
 CC FT 3219. .3399 /note="sequence derived from pMAL-c vector"

XX Key Location/Qualifiers

FH source 1. .3399 /organism="unidentified"

XX Sequence 3399 BP; 1577 A; 509 C; 797 G; 516 T; 0 other;

Query Match 4.0%; Score 97.4; DB 27; Length 3399;

Best Local Similarity 46.3%; Pred. No. 6.4e-08;  
 Matches 320; Conservative 0; Mismatches 371; Indels 0; Gaps 0;

QY 147 AGGAAATCTAGATCTATGATGCTGTGATGGGATTTTGTGATGATGCTGCA 206  
 ||||| | ||||| | | | | | ||||| | | |  
 Db 2202 AGAAAAGAAGAAGTAAACACATGAAGAAGAAGAAAAGTAAACATGAAGAAGA 2261  
 ||||| | ||||| | | | | | ||||| | | |  
 QY 207 AGTTTATTAGACTTAAAGAGAGATCTACITTCAGAGCCAGCAGTCCGCCAGAGAGGC 266  
 ||||| | ||||| | | | | | ||||| | | |  
 Db 2262 AGAAAAGTAAATGATGAAGAAGAAAAGAAAGTAAATATACATGAAGAAGAAAAGAGA 2321  
 ||||| | ||||| | | | | | ||||| | | |  
 QY 267 TGAGCCACACACTGAGCCCGAGGAGCTTCTGTGGAGGCGAACCCGCCAATATCGA 326  
 ||||| | ||||| | | | | | ||||| | | |  
 Db 2322 AGTACACATGAGAAGAAAAGAAAGTAAACATGAAGAAGAAAAGAAAGTAAAC 2381  
 ||||| | ||||| | | | | | ||||| | | |  
 QY 327 AGATGAACAAAAGAACAAATTCAGTCCCTTCTCCATGAAATGGTACAGCGAGACATGT 386  
 ||||| | ||||| | | | | | ||||| | | |  
 Db 2382 ACATGAAGAAGAAAAGAAAGTAAACATGAAGAAGAAAAGAAAGTAAACATCA 2441  
 ||||| | ||||| | | | | | ||||| | | |  
 QY 387 TGAGGGAAGACTTGCACAAAGAGATGGCCACAGGAGAACCCACACACAGAGATGA 446  
 ||||| | ||||| | | | | | ||||| | | |  
 Db 2442 AGAAGAAGAAAAGTAAACATGAAGAAGAAAAGAAAGTAAACATGAAGAAGAAA 2501  
 ||||| | ||||| | | | | | ||||| | | |  
 QY 447 TGAGTTTCTTATGGCGACTGATGATGATGATGATTTTGAGACCCCTGGAACCTGAAGTATC 506  
 ||||| | ||||| | | | | | ||||| | | |  
 Db 2502 AGAAGAAGTAAACATGAGAGAGAAAAGTAAACATGAGAGAGAAAAGTAAAC 2561  
 ||||| | ||||| | | | | | ||||| | | |  
 QY 507 TCATGAAGAAACCGAGCATAGTTACACGTGGAAGAGACAGTTTCCACAGACTGTAATCA 566  
 ||||| | ||||| | | | | | ||||| | | |  
 Db 2562 ACATGAAGAAGAAAAGTAAACATGATGAAGAAGAAAAGAAAGTAAACAAAGTAAAC 2621  
 ||||| | ||||| | | | | | ||||| | | |  
 QY 567 GGATATGAAGAGATGATGCTGAGCAGAGAAAATCCAGATCCAGTGAACCACTAGTAGA 626  
 ||||| | ||||| | | | | | ||||| | | |  
 Db 2622 ACATGAAGAAGAAAAGTAAACATGAGAGAGAAAAGTAAACATGAAGAAGAAA 2681  
 ||||| | ||||| | | | | | ||||| | | |  
 QY 627 AGATGAAGATGGCACCATGATACAGATGATGATACATACCAAGTCTATGAGGACAACGC 686  
 ||||| | ||||| | | | | | ||||| | | |  
 Db 2682 AGAAAAGTAAATACATGAGAGAGAAAAGAAAGTAAAGATGAGGAGAGAGAGAGAGA 2741  
 ||||| | ||||| | | | | | ||||| | | |  
 QY 687 AGTATATGAACCTCTGAGAAAATGAAAGGATAGAAAATCAGAGAACTGCTCCCTCTGA 746  
 ||||| | ||||| | | | | | ||||| | | |  
 Db 2742 AGAAGAAGAGAGGAGAGAGAGAGAGAGAGATGAGGAGAGAGAGAGAGAGAGAGA 2801  
 ||||| | ||||| | | | | | ||||| | | |  
 QY 747 GGATATCTCTAGAGATTCACAGGTAATTGTAGAGAAGTAAAGCATTTTCTCTGTGA 806  
 ||||| | ||||| | | | | | ||||| | | |  
 Db 2802 AGATGAGGAAGGAGAGAGAGAGAGAGAAAATGAGGAGAGAGAGAGAGAGAGAGA 2861  
 ||||| | ||||| | | | | | ||||| | | |  
 QY 807 AGAACAGCAGGAGTACCACAGATCTTAA 837  
 ||||| | ||||| | | | | | ||||| | | |  
 Db 2862 AGAAGAAGA 2892  
 ||||| | ||||| | | | | | ||||| | | |

RESULT 9

E08995 standard; RNA; UNC; 1686 BP.

XX E08995;

AC E08995;

XX d1107332

XX

DT 07-OCT-1997 (Rel. 52, Created)

DT 07-OCT-1997 (Rel. 52, Last updated, Version 1)

XX

DE DNA encoding a protein involved in Leucocytozon phylaxis.

XX

KW JP 1995089995-A/1.

XX

OS unidentified

OC unclassified.

XX

RN [1]

RP 1-1686

RA Kato A., Onaga H., Ueda S.;







exon  
PNGTTLHTVTCREGYAKRQVQVTCVNGNWTVPVKCKOKKCKCSTPPQDLLNGRYVTGNTLY  
YGSVITYTCNGSYSLIGSTTSACLLKRGGRVDWTPRPCKIDKKKPPQKPPQIANGHTTN  
VKDYTYLDYTYTSCNDTTLTLTGPSSKLACSETGSPWPNGETKCEFFICKLPQVANA  
YVEVRKSAATSMOYLHINVKYKGFMLYGETPTNTCHGVWSPAIPCEMKISSPKGDMGP  
INSNEDNSTPSGAECACGSPNPISS"  
10912..11773  
/gene="mCCPH"  
/number=1  
gene  
10912..11994  
/gene="mCCPH"  
gene  
10912..12013  
/gene="sCCPH"  
CDS  
10912..11994  
/gene="mCCPH"  
/note="ORF 04a"  
/codon\_start=1  
/evidence=experimental  
/product="complement control protein homologue"  
/db\_xref="PID:g60324"  
/db\_xref="SWISS-PROT:Q01016"  
translation="MYTLHYICLVLSGVYFVWTLSCPTRNOYVSVKYNLTNYSGPY  
PNGTTLHTVTCREGYAKRQVQVTCVNGNWTVPVKCKOKKCKCSTPPQDLLNGRYVTGNTLY  
YGSVITYTCNGSYSLIGSTTSACLLKRGGRVDWTPRPCKIDKKKPPQKPPQIANGHTTN  
VKDYTYLDYTYTSCNDTTLTLTGPSSKLACSETGSPWPNGETKCEFFICKLPQVANA  
YVEVRKSAATSMOYLHINVKYKGFMLYGETPTNTCHGVWSPAIPCEMKISSPKGDMGP  
INSNEDNSTPSGRICNGCTTSMPTQTITITARTYSTHIYPTGTXYLKPGLVLIIL  
TTSFIIGIILGTGVLHRCRCVMSQ"  
11774..11966  
/gene="mCCPH"  
/number=1  
exon  
11967..12013  
/gene="sCCPH"  
/number=2  
CDS  
complement(11995..12285)  
/note="ORF 05; KFLF1"  
/codon\_start=1  
/db\_xref="PID:g60326"  
/db\_xref="SWISS-PROT:P24912"  
translation="WTEIFLHLAKHLKLOAENNYNSNDPVYELIQLSBEENRNTT  
SLNSIYDDVCYPEDMDYDDCGATRLDASVDYQGLRTCGNDGSHIYEIG"  
12584..15970  
/note="ORF 06; KRF1"  
/codon\_start=1  
/product="major ssDNA-binding protein"  
/db\_xref="PID:g60327"  
/db\_xref="SWISS-PROT:P24910"  
translation="MATKAQPSADNLGRAPVEPCGYIYVYPRKGFPPKEASLGAGN  
NVGASLPLLSLDTVESNFSFNKAVHKRDKMTLLVLRVSAYHREAIVFTNTLFE  
PIFVPGDLILCDARSDFGVTNVPRTDLRDTVDIKDLYAPFYSEDSCFMAVVTYEG  
FKERLFGNLPYITAOGLKVOINGREAVKIPLYDEDLFSKSHLPRYIPSYKYLH  
DSVTSIAQALRIDVESVIRASEKQIQDQYKLAKIVNSKDFSLQSQCDASAFMV  
IDCLAEALAMSGISLFLAOPDQCAVLDTSWPIFETAETEDRKAIQDWMAMSVH  
VYTHLSTNSVLYLTINKQPSKQNVNTYFMQHLAYAADATQRENGEPAFG  
APKFSGGTYTLHLAASFSFPHLIARNYMQFCQHQKSTTANYSVPQVYGTAAAS  
DLCELCQTCASCIHTLFYRLKRPFPVLGSDORDPVVTVGSQYNDLQMLGNFAT  
FREREDVAQNAEAEKTYWOLIONVEKLSMGVTEGTSVSLITDQSLKFRDI  
DNVYDSEVYKMNGLKNNINPREIKITHVHLHYCCNFWQAPCMFLFYKSLVA  
LIQDICIPIANTYDQDNGSGMSEWLKHYQITWTFKSCDLRGLVTSSEKIVH  
TDFCFDLNLSLQGVIPMKQVRLAKALLTPKTIKINRIVFSNMSMETIQSG  
FIKATKRDSYIVTGPYMKFLNSLHKVFNPAKISALYLWHTFSOKKOLPVLPGISRE  
MVELANVETSSKHDDMMVLDIIPITLLTYAKVRLANTILRTCGQTQFYATLQCL  
LPTLQITISATYPHVLLDQSIMSDHYLSSIKDRHALVQVTLKEDIAVQKQRFIVT  
PVLVNYTGINGNTQIFCCNLGTFMGRGVDRLNIPDSTFRNNNSYMMRRRHVF  
TPMVAHLVKKNLNNLTFEVETIRKNQONIFEDKNLNIIDFNVDYVLELVGKSDSCEN  
ITEDDLQFYQYEEIYIMSEIWSRFQIITDGAQWSPVENVTYKLVGCKQCECKFEFVG  
EQLSCVPPQTEEFAPQATLTLAASRRKRTSILSDIDL"  
15977..18016  
/note="ORF 07; KRF2"  
/codon\_start=1  
/product="transport protein"  
/db\_xref="PID:g60328"  
/db\_xref="SWISS-PROT:P24911"  
translation="MAQLAAYSIQVGLTLDVSLITFVDFPAHINWKTVTNTKINK

ITYLQIMPLHNQNNTESTSLVELQLHLHLNKLIVLETSAHLSHDYNNMYFENHISLASP  
CSRHSIYVQFNCCSVKCIINDIIFSRKLSVFCIRSCDALRINHVIDPFLG  
HLRGVSPPLPDTYLSNIPCIYCLNEHMLNQGESLPSLMCMVCKNKHVKCOLNPPI  
QGMFENELQRIIIVESNKQEQNOTCSIDNQRDASLSKLNQHIIFENISAPVLELS  
NLIYSSGAHKKCANVENTSEMAKLLSTEAKMQNKYICKNKELFVRLKLOGEDTIVSHAA  
FCGGIFSSVDDTVTKSLSDCSLAFMKRANYOOLIKKQNELFVRLKLOGEDTIVSHAA  
SAVPLSDRATIVNPDQVLHDAHARKDAYLQVTKDGLSLYCLTQGAVALNTLSMR  
VNGAVYDEIVKKNHFLRQDOFISLDWTHCETDSTVTFENSKYIKNLISYOKLSSEH  
TSSLQYFKLITGLPSONVSFFPLPSNIALAHCLDAGALPHKLLLTETMTWPSLEP  
KDWVSQYKNEYTITSLDNLNIOKEANFFIRELVLSVLYNEVLKNNLYLEADKLIKPGN  
NCVNLSPNOFCGSIYITLSDNLNIOKEANFFIRELVLSVLYNEVLKNNLYLEADKLIKPGN  
18006..20432  
/note="ORF 08; KCRF1"  
/codon\_start=1  
/product="glycoprotein B"  
/db\_xref="PID:g60329"  
/db\_xref="SWISS-PROT:P24905"  
translation="MVPNKHLLLIILSFSTAGQTTPTTAVAKNKTKQAIQYFYRYR  
CSAATTELFERFDLDRCTPSTEDKHVKIGILLVKKNIIVPIYFKVRYRKKITVSRIIF  
NGWTRGVATINKWELSRAPKYEIDIMDKTYOCHNCMOIEVNGMNSYDRGNKKT  
VDLKPDGLTGATIRYISQPKVFAADPGHMLGTYRTVNCNCEIVDMFARSADPYTFV  
TALGDTVEVSPCDVNDSCPNATDSVQIDNLNHTVVDIGNRATSOQHKRIFAHTLD  
YSVSEAVNKSASVCSVFWKSFQRAIQIOTHEHDLTYHFIANEITAGFTVKEPLANFTS  
DYNCLMTINTLTEDKIAIRVNNHTPTAEYQTEGGMILVQPLIALEBEAMELEA  
TTSPTYPAPTSSSRKRAIRSIRDVSAGSENNVLSOIQAYDKLRQISNNVLELEA  
ITWCREOYROMWVYIAKINPTSVMTAIGKPVSKALGDVISTECINQDQSVSI  
HKSLETNNDICYSRPPTVTFVNSSQLFKGQLGARNELLSESVENCHQNAEFTFT  
AKNETYHFNKYVHTVNNISTLDTLALNTFIENIDFRAVELSYSSGERKLANVF  
DLETMFREYNTYAQISGLKDFDQNSQRNNDRIIODFSEILADLSGQIKGVINVASG  
AFSLFGIVTYGILNFKNPLGGMFTLLIGAVIILVILLVRRNTNNSQAPIRMYPDV  
EKSSTVTPMPETIKQILLGMHNMQEQAYKKKEQRAARPSIFRQAETFLKRKSGY  
KQISTEDKIV"  
20507..23536  
/note="Orf 09; KCRF2"  
/codon\_start=1  
/product="DNA polymerase"  
/db\_xref="PID:g60330"  
/db\_xref="SWISS-PROT:P24907"  
translation="MDFNPLYSKRPTDTKPKLHTTRQSCRLVVKPCFRNTEKGV  
SVSFSALPTYFFKGNENKYLLENGSMHLLRRPCKNALLEQSIIFHIYDIYETTYSE  
KQALNDEPFQTDIIPNGTVLKLGLLEGASVGVNFCQSNFYVYKIVPEGNNITLYI  
KOALNEKSPSCAYOTEAVKKILSKRYDPEEHDFVKTVSSLSYVKISDSLVNGCE  
VFETNVDAIRREVINDPSTFCWTKKACPRITNRDSDHTDIEFDGCVYDLEFHADRT  
EWPPNIMSFDIECIEGKFPCKAKNEGDLIIQISCVFHAGALDTRTNNMLLSIGCSA  
WENTYIEFFSEIDMLHGFSLIRDFNVEITGYINSNFDPLYLDRAQTIYNKLDS  
YSRVKIGESYQIDIMPLFMSPGSRKIGLYCVIDSVLVMKLLKFMHVEISEIA  
KLAKIPTRRLVTDGQIIRVSCLLAAAEANVILPVSNVDNADGFGQATVNPIDGFY  
NNAVLVDPSILYPSIIQAHNLCTSLIPHALHNYPHLKSSDIETFMSSGPIHFVK  
KHQASLSRLTLVMSKRAIRKLAECEDLTKIILDKQOLAIVKTCNAVIGTGV  
ASGLLPCISIAETVTLQGRMTLEKSIFFIEMTPDTLQEIPIVPHIVKPEADKFRVIY  
TDSLFCVGVGYDVTYVFGDFLAAFTSEKLFNAPIKLESEKTCQCLLLAKKRYIG  
ILSNDKMLGVDLVRKTACKVQNTSSKILNLKIDPEKVAQAOLLSKTDPDYAFRE  
GLPDGFLAVIDLINESHKRLARTGOVPELFTSELSTSRPISYKTEINPLHTVYKKII  
TRHEPPQVHRIPYVFGVTKTSCISLNAEDPTVYVQNNIPIAVDLIDFKLLHGVANI  
IQCLFKDSKSTVSVLYNEVSTPVLSFELLYDHSVKA"  
23573..24796  
/note="ORF 10; KCRF3"  
/codon\_start=1  
/db\_xref="PID:g60331"  
/db\_xref="SWISS-PROT:P24913"  
translation="MEPSISFVGRWTVTVNSGLFSISNHEEIPATPFPARLVHLPLF  
SIESIDORIISHLVPTLIQLOQYNTNSPAVFSYDREQACKPVYLPVFSDFINPLI  
VFEVGGPIILIKNELKHFVITSPRADISFIPPEDDQTSKSLTLAGFQKTEINEI  
NICGPTMTKEAGSVMLFCRCKIPPEYNTSFDAKDSGLILETLIIQEVGADTYMTV  
HLCSPKPAHIEVQVHLVNLSPLELVFKHALPWEAPSGSVLPIEADKLIKPGN  
SVEICFSFVNRGLSQDPALFVASSNHTTVVVKPQIWPITPLSITVITVNSRRII  
FIKRGFVAVAVPCFFHLKAPQDCEERVILDRENSSIHWSVLIKPGAGGPIVHVH  
VALKEINLTPEPMF"  
24799..26016  
/note="ORF 11; KCRF4; similarity to Lf2 of EBV Raj1  
insertion"  
/codon\_start=1

/db\_xref="PID:g60332"  
/db\_xref="SWISS-PROT:P24914"  
/translation="METPKHIRGVWRDVSVLKHKREKFNYSFWNCSEVHEHFSITNNR  
ETHLPDLDLLKRAILFKQIPSPKFSRSGPPGTVTSLVYVGHSGNLIKTKP  
MVSCEDELIFKIFALACTIPPGSMELFIPFLKLDGLYLCLLEDYTKRSTMS  
CHMGTYLASETPQVFLKGGPVNLKHEPMPYLMQAOKPFDKKMARVHTVQNEVCVN  
SIYRGNHVKAIQKDSIEDINFLDTVVGLTNTNRAVAFENPYSFCPWKKRQSIIP  
IYDGPCRIPAGRAPRYNNYTSIYPNATAIITNSDCADFHISDCWPKGTAC  
IKVNTSIVTSISSGTHGEAVFILAPKFECKIMSKTHVQKLSSAICLPGNVTINS  
NKPKLADLSTVK"  
complement(26045..26554)  
/note="ORF 12; KCLF1"  
/codon\_start=1  
/db\_xref="PID:g60333"  
/db\_xref="SWISS-PROT:P24915"  
/translation="MSSIQKCLICNIGEEELLQACDCPSVHHTCLOSHTOCFKSS  
HCTFEKRYKINWNCNKKSCPSVLEQANWVLCVCTICILICILIDILITIKLW  
QSSVUCYEYINTFYFLVLCGTFSAFYLAAYDIFFEFHSCLSFINWKKISQSYCE  
ASKNALKIL"  
complement(26936..27391)  
/note="ORF 13; KCLF2"  
/codon\_start=1  
/db\_xref="PID:g60334"  
/db\_xref="SWISS-PROT:P24916"  
/translation="MTFRMTSLVLLILLISIDICIVKSEITSQTPRCCLAANNSFPRSV  
VTLNRNNTSSKRADYNNRSTPWLHNRNEDQDRYPSVWEAKCRYLGCYNADGNV  
DYHNSVPIQOEILVYRKGHPCPNRFRLEKMLVYVGCVCVPIVHND"  
complement(27674..28423)  
/gene="IE-G"  
complement(27674..28423)  
/gene="IE-G"  
/note="ORF 14"  
/codon\_start=1  
/evidence=experimental  
/db\_xref="PID:g60335"  
/db\_xref="SWISS-PROT:Q00997"  
/translation="MALDLRLNKLHTANFSFRIMIWMICLALPTDSKPISTEAPIL  
NITQSPNISSEPLESEPLKNCITFDLLQWRGENASIKDMLMTQREVEHGRM  
TTLSPRSEKVEEQLOPRNLLPTANGPVHVKYRLYNRLNEAPKAGADYNGKIQFD  
DPPLPYGAYNDGVLWNINGKHVRFDLSLWRIKRSRGTWCINKTSEKAAILQKL  
LKAEEKRTTSKRITLEEIQIKELEKTSTSP"  
29231..29596  
/note="ORF 15; CD59 homologue"  
/codon\_start=1  
/db\_xref="PID:g60336"  
/db\_xref="SWISS-PROT:Q00996"  
/translation="MYILFTLVTFVCKPIHSLQCYNGSHSTMQCTTSTCTSNLDS  
CLIAKAGSGVYRCKWDDCSFKRISNQLSETQLKYHCKKNCNWKNGIENIKRTIS  
DKALLLALFLVTANFPL"  
30322..30804  
/note="ORF 16"  
/codon\_start=1  
/db\_xref="PID:g60337"  
/db\_xref="SWISS-PROT:Q01001"  
/translation="MSSIKFQNIENILKCCODNRHSDSVVRVAVSHVHOYNKFEAL  
MPDESICVHDRKFTGEALLTETHTNGKGVAMLSFAVLQIIDEYKCVATSM  
SSYISRSVGANWFIENGKESLVEFCNSIMPQNPFLVLPVAVLAGLVLMQTLIK  
complement(30837..32264)  
/note="ORF 17; similarity to other herpesviruses (HV)"  
/codon\_start=1  
/db\_xref="PID:g60339"  
/db\_xref="SWISS-PROT:Q01002"  
/translation="MSIVYAGFVDVAVYKVPDVLVLLMDVSKCLPLTKPIPLANIE  
HLPESTIGHTGLYATGVGCVGIHSEKFLHTENLFSNCSVAQATSKFLPYQPLL  
EMLHTWLPALSLSCPTAQAANNTMFQHSVLSICALGRGRGVAYTSMNLEALISQFC  
SISOAEVNIQDSKNVINSPLKPVFNIDPHILMAKADAGIKDRLOLLKTDKGA  
KIKKTLAKASIGAPVTEISEDNMQGVIPQSGDDLSISVPKFSLSMLNNDLNFK  
QHRPACPFQYSPGAGYNYELYPQYSGNIGYMLPSGSIYVPMFSPRPNKRRKE  
DFDCVFFGESLSKDVNLTKNISLODLDLKDLOAAINQPNRYPPHFSNPYSIDLP  
CHASYFRVAPYKAPQDQHLLOPLACVQQAQVQVQPNYAPPTPEGASNEAPKVSQEPV  
HIDASFAODPYSKLQKMCDELLNK"  
32257..33027

CDS

CDS

gene

CDS

CDS

/note="ORF 18; similarity ((sim.) to other HV"  
/codon\_start=1  
/db\_xref="PID:g60338"  
/db\_xref="SWISS-PROT:Q01003"  
/translation="MDILGRFVRESKMSPLKATIMEKVLRGOSLNSFHSEERLFLHL  
VICKWDFCLNVIKLESTINTDRNEVLSRKVPVEIWKIMYDACKSIGVENTMLID  
DSRGQWLHLNLSNIDLQMSQFIFSKLGIKHFVAKISQNIATDGNLYLNLGSLVPLR  
LILILQCLIFWGKEQEAWRFTGKIEMLYLILITGHLLIQTKTFLILQAASATGVCGL  
EIIIGDRLSYLGIHTYNDLOHIPSLLDLDFIENNEY"  
complement(33024..34655)  
/note="ORF 19; sim. to other HV"  
/codon\_start=1  
/db\_xref="PID:g60340"  
/db\_xref="SWISS-PROT:Q01004"  
/translation="MMKLEKYLILRONPSVFLNGTAFWTPHPONILHIDRNSLRETKK  
NASLYTRLNLNLETOIKKAMINVELDKMDHVKRSSIITRLEIEIENNVKFDQSP  
QLPSPKPLSPPTQPTNTNVTYITVAPDGPFTVESNKIELVSLYTNQOQM  
LPSYGPWYSLTDIAMQRRVFPKELRGTNYQNSTSLKMHAVLTITSSATDGFYSV  
RHSDTSALVILNAVFLCKTSAPIPVYVEELNLEAKLGMFVDELKNTGGNGRFF  
SPOVNEATSIAPPNKDTKYSOHFSSKIYSLLEASGLLSTKSHETNPKTVDIYIT  
TEIFGEDIPPMASFQWNLRVGIIAIEFVITYLLETSQISIHSTRRLNLSLLGSK  
FKASSTGLNLQIVIKGQVFSFLKNYIVPTLTHKNKNTVSFLFPGVTLLALESLAT  
AVDKPFINLTGNRFODFEIINQKFTKDPVSLMAAQATLRFKVEHGLSILNLSPT  
TTEATIRIQFGGEDDYDTLYFIVLGLCLPAWAAV"  
complement(34285..35196)  
/note="ORF 20; sim. to other HV"  
/codon\_start=1  
/db\_xref="PID:g60342"  
/db\_xref="SWISS-PROT:Q01005"  
/translation="MLSVIKQDKVEVLAHLPNKRKIAGNAKHALETYKKLAKYTVSASI  
EKFLSISHPCLPRAKTRLFEVSLGNRIADCVMLTSCGETRICYVLELTKMTSNL  
ISDIRKSQSGQLQADVFNHYNAPLGRQAWTVLPILIFKSQTKLTLKHLETPEF  
PVNLTHSEKSLCFLWSRADVEIRKKIHLAPKPRIFKWDSDLDSTDEHSAYROKL  
IERNKKCTFLQNTQSKFRDTRNKNSDQLRQANARPCCKKQHNKRLNRKHHGG  
KYSRLTSTTSFSEAAFSNYPVSTHKL"  
35195..36778  
/gene="TK"  
35195..36778  
/note="TK"  
/note="ORF 21"  
/codon\_start=1  
/product="thymidine kinase"  
/db\_xref="PID:g60341"  
/db\_xref="SWISS-PROT:P21293"  
/translation="MTGRQPPKKNDIYDPRKQPPKNGSYNDYPTSTKTSTNNK  
RKDSNYPPTRETFEPDLAADPIYSVPPSRVPKLVKYNLNPITASNSVSELL  
SLHDETQVECEVAPLIAKSPDITVYKMFVRPKHTLTLEKQKQFTKKKGSFVK  
IGSNMLEGESLKSKLHNSKSPDEPDGLVHVPHLLYPPKHQDPVPAFFILEGSI  
GVGKTLTLLSMNGILGKGNVLAHFEPIAWDVSNSLEEVYKLTLPKRVGRTNSAK  
LLACQLKFAFASPLALATATDRJSSPKNSLLSDMWYMDRHPLSATVFPYMHFQNGF  
LSFSLHILQWSFKASRGDNIILLNLSQENLARKVRKRNKEKSVSIEHRLINCY  
HAYICAWLLVQNFTEPEIVEVCFNAKHITDLSKSPSLAKHVSIEDMLKSSIFNTWI  
EMTKAHRDSTLMECLLTFCKEKLVQLIHVNVSPDTDDIPGLWASITYTSIRNSAIK  
PNRWVLALEDLARTENSQ"  
36775..38928  
/note="ORF 22"  
/codon\_start=1  
/product="glycoprotein H"  
/db\_xref="PID:g60343"  
/db\_xref="SWISS-PROT:PI6492"  
/translation="MTILQLFLVILNEALCDYQPKPRINKPPAEERLKLNGYNT  
TLIEFDGQVSFNLMWTKIEIHPHDELIELREANVTEPLVNTLKRSDTYRPEYNV  
HYTEGNSVACALPYMSYTIQDQWNTGTGLGNFGIPSKTVLNEEFYQYVYTNRO  
FIPYATVNLCLIGATPAVPTISCHITNPLVSVSEFKEDSLTILFGHSHVLPK  
GHVYNDISGASNDVSLVISTYDILFGKHVESFKDIAKVFREIETPLPTIKNLQ  
DEMTLEIRGCNNINIVPKTLFAFAKVAHFLVDSLTQQHLLNLCFANYMSEL  
EFLRKMESCFEEDFFPYTETLAASQALNPKHVITISLQSHQKILNCSLRSER  
HSKHVSVAISMTIIDLHLLHFGOKDLDVYETPCVLSLRFDTREKLTPEPQ  
DLLPYVLSSSFTLEIGHLYTAYSYTMTLSERKMLDAYIVLNDINHKNLTETPKQ  
SSIIAQKENLNTNGFPOTLHMHHHTSLEIPIKICKLSLTDILISLKNITYVIS  
TPVNSKIIDYSEVFLKTSMTISAVNDCCKPYQGSAAHQPVIITVNTVPRGCGPIC  
SVLSVSDSQGQFMVITDITYQENLEFSEPFEGDGNLHIHYLILMNNNGTVEIC  
GAYRARLVNFIIVIMVFIIVGLVLYLKYLFVLT"

|     |  |                       |
|-----|--|-----------------------|
| CDS | complement(38921..39682)   | /db_xref="PID:g60348" |
|     | /note="ORF 23; sim. to EBV BTRF1"  |                       |
|     | /codon_start=1   |                       |
|     | /db_xref="PID:g60344"  |                       |
|     | /db_xref="SWISS-PROT:Q01006"   |                       |
|     | /translation="MYMQLRSTTKKKVYKINEANNMAVFEKNKPPSSIKRAELKTVLIPRIKNGKIGCGLEKTAAYIKAKAHPVLGKQSPRYHTINWTKSONALILPSPMDIKNTQHLKSLDLYGLKSNQFPAFGYVLEPMTSKEMLSQFDLFTKRTEDRIIVLVTNICEHREHVHVRASVTQVFQVDIVKEVFFLPAPKDKSNSINFAASVVEIICNGPFSKVIQVYNAYMDIKHTTWNLIKIVALLTT"  |                       |
|     | complement(39692..41887)   |                       |
|     | /note="ORF 24; sim. to other HV"   |                       |
|     | /codon_start=1   |                       |
|     | /db_xref="PID:g60345"  |                       |
| CDS | /db_xref="SWISS-PROT:Q01007"   |                       |
|     | /translation="MSMFLPVNIPPAKGLKVVQTRDCNENIQLKIFCCLATNOVVSLSLDFNQLDCIMGHESTFEYTCRAVRRLLLGNSWYFFDITNESATGTRGPMYNGPLIIINTDSTYKLTNCSNKLYLDIVSLETDMPHEPLATRAYIFPDLEQTPIDIMCMFKIKIRYVTSLEDECEYELFATVPFVFNCKKNYLRLVSALKTLPTSLVATPNPPDQLEICRYVTSLEDECEYELFATVPFVFNCKKNYLRLVSALKTLPTSLVATPNPPDQLEFKFSLNMQVMPNFAETIKRDPGSRVSLVLLPDGVKVVYPPSLPYRVVTMCLASVAAVQSTITSDKODITASALVSI FRKINYPADKKDKDVLNSKVALDIFERNYIAQKTVAEDEEFTYRPIHRLGINFNKVNVENTNVLTKIFTHSVPTWYKSLMDIPLRKNFVKYYSKEPETSVEYFCENCMGAAININISGALLTLFLHMGNNKVPKINILNIPVLSWNSTLDLHGLNLSIVTRGRNDVTFNFPSSVLTFLHMGNNKVPKINILNIPVLSLTSHTVREKSPPIISNSARLSLKNLNSVLESRRNSQIOTLHKRMLCELVACSFRLDSHTVRLAQKGFDFSKKTIISHAKNKHDCALLGYKKNMIPKVLNSENKKTRLDEHGRNANLAFISATGPHVPKIRAKLIKHLVLTGLHWRRTYKHYVNTAQ"   |                       |
|     | 41893..46008   |                       |
|     | /note="ORF 25"   |                       |
|     | /codon_start=1   |                       |
|     | /product="major capsid protein"  |                       |
|     | /db_xref="PID:g60346"  |                       |
|     | /db_xref="SWISS-PROT:Q00999"   |                       |
|     | /translation="MEVENRPYPYMSDANLLOQIKESSAEGLFKPSLLIGKDVRESGVKFEALLGYVNTAQVFXFLETSLAVSCVNEFEKOLKRMTDGKIQKINVPYTIAGDGRPKQKQFIIMKATKHHIGAGIELSTODLELLFSLKETPLDIDYEVGAVTITISA LQFGDALREGLDITVTKLRHAPPLFLIKTLADPTYTERGLKKNYKSLISMFKTHLVNNSFLPHSRQYVLGILTEMIGAVCKETYFKGISYSTANGOPISGVLET TDKVMRKLNVIMGADNSIMPAAAYNVVGRGNLVAISYKAMRNFDHFMKSLVDNPTSNLDNDVDTFESTGISQIKTPISTSVVMGNKLLALESLOMYNETQPLPLNRRKHVYVYPPVGLHSPKYSMSVSGTENVLHQSEAVINVKNTLQCFNQNALKSIC HPRMNSPICARAGAFPDVHNLNIGKSEDAHTMNLQYIVDYDNDKHHVAVHSLAQKSMTHBEVHLTHETLREVPFFDYAERHOGAAVOYRATHRLNSGNLPPPLA PNYFCQRCYGEVASLGHVIDSTMEITQETAFDPAYPLLCYIVESMIHGOBEKV MNIPALCIQTYWNNSGRLAFNSFYMLKFICTHMGNGHISKDAYSCYRKIYIGELTAIEQSLYRLAGHENAVENIGOLNAILDKLLPPEAYNDITLNLKSSRHPYVKIGMIEYDDNDQNCINIREKEDLVGNMNVYIQRRNTDHSRYVLDVGELENYNSVL EXLIFYLPLVCINGHVGMDVFENVALTLPVFPVAFANODADIDHLENGTLR DVLVASEIRPTVGMIRRLCTSPFFITQAAKIKTRDPGQINIVHTIDGKYVHOTVYLVNGFAAFATADKSRDAHGLFTYVPVFNKLYCDPMVAATLHPVIAEFTTEIPSONNAV FNLNPLRIAIEBWEHKSPMSVSVSTQSPGLSPTMIAMHLKLSPVFSICQSRHKIHP GALTAVRDEVVAEHMYSKASTSQVGPVHRKERSDVAFFDINHELASLDPA LGYSTVIPAHAALITDMGICHCODIFAMPPESEAYSNQOLNEIYKQIGSDRYGMPLRDPRMGNRRVTLPLGSHGQATECIVIMPVADITYFQSSNPSRASCVVSCDPA YNNESAERKLDHSLDDPDCYEFRTSNIPWASQIGSLGDVFEFNQHRQMAQGLTYSPOKQFFNKEAILKNKRLFTLVTEYVNRULTGAPATNTIDFYVYVINGTDVFLPQPCQFLQEA YPTLSASHRALLIDYMSHKTTHAPVHVNQYLVEEVAPMKRLKLVGNKNTYV" |                       |
|     | 46024..46938   |                       |
|     | /note="ORF 26; sim. to other HV"   |                       |
|     | /codon_start=1   |                       |
|     | /db_xref="PID:g60347"  |                       |
| CDS | /db_xref="SWISS-PROT:Q01008"   |                       |
|     | /translation="MLTDKTIIVLSLSRLFADEITTKLOKIGISILPLQDPHKLQSLD LGLNAVCSRDVPDYPVHMPSYLSKCTLAILEEVNPNLILTRLDSPSETVQIKNYEPM FQWDGFSNLTVIPVYEGROQATVLTESNGFDLVFPSPVSDLAQALIGKLLNLYSR LVESDEPTNIEEWNMTTNTWTHGRHIVLDINHNHNEALKSLDLDLAVTYCILSALIP RACLRLVTLIHRHDDHLLDVFGRVYIPREYVEIDANALSIGDDITRMTTPTIYLOSIS SIFNLGAKLHLSYASETQATCWISYC"  |                       |
|     | 46945..47787   |                       |
|     | /note="ORF 27; prosite motif for cytosine-specific methylases"   |                       |
|     | /codon_start=1   |                       |
| CDS | complement(48089..49231)   |                       |
|     | /number=2  |                       |
|     | /note="ORF 30; sim. to other HV"   |                       |
|     | /codon_start=1   |                       |
|     | /db_xref="PID:g60350"  |                       |
|     | /db_xref="SWISS-PROT:Q01010"   |                       |
|     | /translation="MSKPCISEDFKQCOAFKAPLPILVAEISKSLANIDFVDFTASQ QVDQGLFLLDLIGTECFEKLCPHTSANLTN"  |                       |
|     | 49443..50069   |                       |
|     | /note="ORF 31; sim. to other HV"   |                       |
|     | /codon_start=1   |                       |
| CDS | /db_xref="PID:g60351"  |                       |
|     | /db_xref="SWISS-PROT:Q01011"   |                       |
|     | /translation="MSTHQCFQFNKLTNVVQYHKVFCYQCCQCKKXIHCDGSKTCIL INTGESVLTCTGNCFINNFQENTFLKQDNIKESQDFDTLFAILLESKLQDLNEF IKSYLEVSKAILDNNLSKPEICDLIRCTFHQCHIFGEIQYGVDIICSMYTHIILS IYSPRTVGNLLFRCTNKKYDYLKRNRETWMSTLIITGDSQDOQIKS"   |                       |
|     | 50015..51340   |                       |
|     | /note="ORF 32; sim. to other HV"   |                       |
|     | /codon_start=1   |                       |
|     | /db_xref="PID:g60352"  |                       |
|     | /db_xref="SWISS-PROT:Q01021"   |                       |
|     | /translation="MDVHFDNRFLSRPDKVIVHLLLPQTFEESHIIIPDPNLTWFSQ TRFHLLDPPTRYVKINGKLYFLSPGLPSKACPGIANSVLINENMNPDMNTVLKI LLENNTYIKFHHMELEFPYIAPNSLQDTSIEEPFNKVDHIDAILNTSPVNTNLYK LGVLANLQSRQSPSRHFALEDPGKRYGEQPSLQTDKAMPKSVYVKHKSILNS EPTVTCVKHIFTKKSIFYFCSTPTMSNQCTSVLDTISVQELAHVSTCVLDAEAKFL KHQHRFVNTLHCVCKSKNYAIEQHPVLIQIDEETASSIKDHFETCTCVLESTVSEAS AWNRATPARYLNKPKAFWIDYIRLWEEGVHPLGKSVPELKEDEVCEEKWHLLSLNKEP THICRNGCGTGLVDLSNLNAILPGFVVIKGNITPEDILFVGARG"  |                       |
|     | 51333..52325   |                       |
|     | /note="ORF 33; sim. to other HV"   |                       |
|     | /codon_start=1   |                       |
|     | /db_xref="PID:g60353"  |                       |
| CDS | /db_xref="SWISS-PROT:Q01022"   |                       |
|     | /translation="MDDFRKLRFLNNECLWNVAVCTSTKTYCATVAVSPFPKPI SPOGVDPKHINVTLLILKPKKSHPHYITVYINDLAVDCCSTEILQVRKVPVCSHFSLIY FGPLIAPPNVQIIPANLSIKAKSKSLTKNQVIETKSHIPRPLPDGKYKASLIGACCA WYSEGAIFQHFLLSTDYMSLCPAPKFEFPLSLRILSLLTRCDDLSCVPCYKGEIHNVCAS WYSEGAIFQHFLLSTDYMSLCPAPKFEFPLSLRILSLLTRCDDLSCVPCYKGEIHNVCAS   |                       |
|     | 52326..53223   |                       |
|     | /note="ORF 34; sim. to other HV"   |                       |
|     | /codon_start=1   |                       |
|     | /db_xref="PID:g60354"  |                       |
|     | /db_xref="SWISS-PROT:Q01023"   |                       |
|     | /translation="MDDFRKLRFLNNECLWNVAVCTSTKTYCATVAVSPFPKPI SPOGVDPKHINVTLLILKPKKSHPHYITVYINDLAVDCCSTEILQVRKVPVCSHFSLIY FGPLIAPPNVQIIPANLSIKAKSKSLTKNQVIETKSHIPRPLPDGKYKASLIGACCA   |                       |

exon  
GTTDDCGKSNCSCTCTALKKDIIVPITGHRNLLSLFLDATQHNTITSKFFSPQ  
TPTTVNVFCGLDGETVECTCEANLLMFSDFSIQMIYNCQIMKRCFLRSC"  
complement(52315..53223)  
/number=1  
CDS  
53222..54172  
/note="ORF 34; sim. to other HV"  
/codon\_start=1  
/db\_xref="PID:960354"  
/db\_xref="SWISS-PROT:Q01023"  
/translation="MENVELAVKGNPELKKRYSGQLDIAIMSENIPDOFKLIETPT  
NSFLISNMPEDSRPHWTQIOKCLDFSNLHLPKLNKIKCTGDHREDTYYKLEAV  
PPYLVDSEENKLAQINKNLSIYSAIEMLADPKNQGGLPIDPUPYIWLIFYGKSF  
CASSDCIYFYKVPGLMPLPPHMYRDPKNSISFSHVQYVKNLYEEYSEPINLEIV  
PFDNCRIEAVELQIDLPVAYLNLCLLCTLRQNMASRSGDGMCGYIIVLGGE  
KYITTIISKRCTVSGDCLIVSYNLSLLMQNNEYEQ"  
54159..54611  
/note="ORF 35; sim. to other HV"  
/codon\_start=1  
/db\_xref="PID:960356"  
/db\_xref="SWISS-PROT:Q01024"  
/translation="MNSNKEFLYSAFEIENKASVSFDRFGKSGCIFLHOLDHTK  
KSLIKHENLKROKSTEGMLQAVDTSTQEKRELSLLKAFNRHKLTAADLQDKILELK  
EDIHEIESLNGQFSSQSEENSETSIDPTIMQWRIEALRPVPSATF"  
54499..55794  
/note="ORF 36"  
/codon\_start=1  
/product="tyrosine-specific protein kinase"  
/db\_xref="PID:960357"  
/db\_xref="SWISS-PROT:Q01015"  
/translation="WASPHRKKRIQVQLPSLTQSCNGELRCPCEQLQPHFKIPILS  
QLITHEKVCSEI SAHKVFKLPSWFOEHLKVPQEDAVLGGSGFGVKPYKSVTCA  
KYTPDVFYHELIACNLAAQIAIRNSTRAIVLWMTGAYIPCKILFPRYAGSLVDF  
KQWDKISAKLAKPEKSLVNAVNFLENDEVGIIHSDISTSNILVAGHPDPGFFLADL  
GVASHSGNHQTELCVKNSRGKILYNMSTRIFFLCKDPVPAQYFICRYLLACKLI  
NSETJHQFFIVKILAQINDMASLFYTMLECEVKMLDKIDKIPREFYKIGPDDESH  
KAYFQFVPRVLEMLSRMLDTKLSIGDSFGNTTEKNVLEVDQDKILFASWCTHLK  
GFLKGGQLNLNKLRAQYELVFRFLKDFYSLGRQ"  
55794..57245  
/note="ORF 37"  
/codon\_start=1  
/product="alkaline exonuclease"  
/db\_xref="PID:960358"  
/db\_xref="SWISS-PROT:Q01013"  
/translation="WDLSEESPINEIGNWMDTQDTQLCSSFSHFLKHPKVQHFIS  
TYSELVKMPTIRVYFYLFKIGGFIGNEKIGTIFSKNVCNNAAGVPRKADVTKA  
CEKMLROOSEICLIEETRGQYLNLSLDALDGTISSKRYMATKQNSKKIPEP  
WPKNIDYVAGLAFGRCEEYKIVLNELICTPKQASCDCGFGQSPDGLGIFVSLD  
YCTNVTNKNLNLVFPDTEVEYIKSRKFLDKSECDTLYKKYKELYSNPCVKTILK  
FIFSVKPAIEFPVPSGRLPSESDYLLAYDEENLRPTKKRKLNAHEMIKKCIENSY  
AGSQYIILSDPAENQITVSKFKAGIFMNPFRHYFYQVALQHRVYVQSYIGLSESPK  
SLGTQKNFIVSSFFRRHFSDDPVYGVGKKQLEKTVIEPVFIITPVYIPRSALLETI  
SQAVNFEESAKEAFETYPWAPCALFANGDLTP"  
57200..57400  
/note="ORF 38; sim. to other HV"  
/codon\_start=1  
/db\_xref="PID:960359"  
/db\_xref="SWISS-PROT:Q01025"  
/translation="MGTCLSCVRRPNPVDTEGKVINVTDDEEMSETEIMLACPODK  
KLCKPKESMYKTKHKNKHGI"  
complement(57425..58525)  
/note="ORF 39"  
/codon\_start=1  
/product="integral membrane protein"  
/db\_xref="PID:960360"  
/db\_xref="SWISS-PROT:Q01017"  
/translation="WKAASRSTEMLRTHWIIQLLVIVFIMFIMSAILPIAASVEGLGFP  
CYFPLNVDYSLNLFIRNAKRLPTFLFEAPLEFYITWSVLVDIASIYVVGALA  
ILQAKKTHLTSMTIQTWINLVGSHTMLFIGIARMWTLOLFIRHVLVSKRVMLAAFIYF  
LHFCLSYMHLSNPSKWSVLLMEQHPKQSLSTIDYKPLCVNNVLSLALLE  
MLVFSLGFMMAGNSGYILVSDTVLASINLYFVLTFWYMTMEMFLODYLKLOFGFYL  
GVFSGSLILLPLVAYEAVFSANLHKTVANMIAMPVCVIAAMPRLFRYSQQVRKP  
ENSYPLPRFKRKRQKQDQQLIMVETSEEL"  
58610..59962

CDS  
60095..60580  
/note="ORF 41"  
/codon\_start=1  
/db\_xref="PID:960362"  
/db\_xref="SWISS-PROT:Q01027"  
/translation="MMTSEASLFQTPQHGVSLMTWELKRLMSQPIQTCMNNIRKPTIL  
MSTQDVALEAPVLEKACSGRYLKAYLYSQSQAIVKPLLENVQWYPYLLN  
NSTIEDLEIHKCEGLYTCSSNSEPLKQTHLEDVFKIVDHSFLVAKTHIHT  
VIPFILLVNVCEKWNGLPEFFRALHSAVYREFTGIAPIFYLPFGGCEATPPDI  
YCGFFPVLNCKGPEKILLCDLPRLHCPOILLISIPGNHADDLLCOPDIPSPIKI  
WPNGIDINSKFCLEETSEAFUNFKHIIVEVDFLCTICYIMGCKSEIYNIMKFGSN  
NUNSTLQTYNNFISTIFKWATQSRFWNTAMTFKYLVSAYASEIPKVSYKLVKQYE  
"  
60095..60580  
/note="ORF 41"  
/codon\_start=1  
/db\_xref="PID:960362"  
/db\_xref="SWISS-PROT:Q01027"  
/translation="MMTSEASLFQTPQHGVSLMTWELKRLMSQPIQTCMNNIRKPTIL  
MSTQDVALEAPVLEKACSGRYLKAYLYSQSQAIVKPLLENVQWYPYLLN  
NSTIEDLEIHKCEGLYTCSSNSEPLKQTHLEDVFKIVDHSFLVAKTHIHT  
VIPFILLVNVCEKWNGLPEFFRALHSAVYREFTGIAPIFYLPFGGCEATPPDI  
YCGFFPVLNCKGPEKILLCDLPRLHCPOILLISIPGNHADDLLCOPDIPSPIKI  
WPNGIDINSKFCLEETSEAFUNFKHIIVEVDFLCTICYIMGCKSEIYNIMKFGSN  
NUNSTLQTYNNFISTIFKWATQSRFWNTAMTFKYLVSAYASEIPKVSYKLVKQYE  
"  
complement(60572..61369)  
/note="ORF 42; sim. to other HV"  
/codon\_start=1  
/db\_xref="PID:960363"  
/db\_xref="SWISS-PROT:Q01028"  
/translation="MEVVARVLGVNKNMACQEQPIVPRVLVKNKNKNCVIAANTP  
PCFNDVLMVEQLKTHIKSRMFSGVFVMACFVENEVDMSINPPHYFSRSLFY  
NCPNHILLMEQGLSKMLKNLNTSSSLLSAIVERAYVLTMSKRCPPATFLLHGKTLIA  
STSYSPYGNAPVESYPLLMFKLYKCIEDGDPVSKGLKPLIYLSWKMDTQYDAPS  
KOLSEKCVNLFVNTVFKHLOHKEVLKFKVCVCTSPSRNLSQ"  
complement(61335..63026)  
/note="ORF 43; sim. to other HV"  
/codon\_start=1  
/db\_xref="PID:960365"  
/db\_xref="SWISS-PROT:Q01029"  
/translation="MNSGLSSALIPVHPKISVHLEILOGKAYAVKGQTLHSLRNP  
GVFSROLFIHYKVLSSCTYDVISDWHKIEDNIKTRWKNENFTSDSFKSTYQSWA  
QTMRTMIDQVLNNIYQIIHKSITISYERYDWMALGVVVKQLDEKLIQIQSQA  
SAASKIAAGNDRVMRTILSALEYVTDMLETLSIYIPDYSEVTIDYRSKDLQFVGY  
KCKHIAVEVITRPIAKNIFEDSPVORFQINSCYRTEHAKLCLNTAPKAICGG  
ASTQNYKDIILSHLEONSKSDPKRELLSLLVLAENKTVSGVDYVEEFTDVSQNI  
VKNLFGQSESTQGLKRVSNVFNKLTQSQNEQFDTINKLEKRELYLGLKINIE  
TQLLHTQDEKQSPDNVLVSTFCSLQTSKLSMTSSDVPKGAALNLSFSQTV  
PPFRELNDLTLWESEIHFETKALTPVDPQQLYKVTQDTVTILIGFTYSITKL  
HOMELNDIFVMSFNDIASIYVSTSRVAVIADIGNKVPSETDSSSSSVTGS"  
63001..65346  
/note="ORF 44"  
/codon\_start=1  
/product="helicase"  
/db\_xref="PID:960364"  
/db\_xref="SWISS-PROT:Q01014"  
/translation="MAELSPFELNMTSDAKVRIIVEKIRKLSNITRPREMTLYNDQ  
FDEQCGILLPTCTIVITGTAGAKSTISALYQNLCLITGATIVASQNLGRCLKT  
YCTPIFNAFGSKSHINILRSPVRLPTDIEQINFELCKWPIFTSIYFEFSKKN  
LQGYSSISLAFAFNLMKMTTNTWIVIDEAGTSSHLTAVPFCYFWFYNSWLNTP  
LYRSGAPCIVCGVSPOTDAFNSTYHQKYNIMECDNILSFIIGNKVVSYSITL  
NNWALFINNKRCOTDPEFHLLKLTLEYSLAISPATMEYIDRFVVPKQILNPLFQHT  
RFLSHAENVKYSLSLHTALVTGTVNVSAGKLFCTPIVCEVFTKAFKSHVNLPSLT  
ATEWLSKLNHRJNSYQFIDQDMTAHTETDTSTVYLYTKVYNTYISLNGTKKC  
VCGYGYTKNFKKILSESEFIDSHANDQPEFVSFCLTYLNSLYNFHYNVTKEKES  
YTLAKDIFISRLKLTATWFSNKFLDREFTSFTINMLVRONIEFTSNRGLGLEVAS  
TVESYKQGYTFELPVNFGRSQTVISKDLODKMPSIVQDSSGCFIACLEKNVKNMET  
LDGGRSFHLCSAGDYGISSKRLAMTVIIRKQSLDKVAICFSPHKKIKVSHIYVAISRA  
TPNPHVMPDNCNPLKLVNDTQISSQHIKALNPNPTLVY"  
complement(65384..66157)  
/note="ORF 45; sim. to other HV"  
/codon\_start=1  
/db\_xref="PID:960366"  
/db\_xref="SWISS-PROT:Q01030"  
/translation="MAMFLDTKAVSKLKKGRKRLWPLPGAPREKQTNVFKFTDGDND



```

/translation="MEDIEGGSSDDDFDSSOSSSEEDSDTSPQIMKSDVMTASPP
STPEPDSASTSNKRQSPRTWEHOSPLSRVSRVSPRFGRRPRISSTSGR
SCTSWADRVREAAQRRSPRKPYPHNRNGPLRNGPPRAPPLKLLFISILPKSGS
EPLFLPVLPCQEAETKNDKYLAMAOQRAMHDVPLSSKOLTANLLPVREKPLLSIV
RTPPNYVWVMKREKTIASANDTVAFLDSESCWGQQYLKNDFISFENGRDILFIS
SALSQVHKMKPCHCLMOTPDQHDHVKQVCYLKSNRNLDAVRYLQTSVIKSP
IVLLAYAVGTPAAICTKNETQILYSHCMRILREYRPGDVNMLIHESLTOHLNCPSS
TCAYTTRAVGTRANTTGLFPLTQ"
<78291..78311
/number=1
/number=1
78395..>79627
/number=2
complement(79986..81059)
/notes="ORF 58: BELF5; sim. to EBV BMRF2"
/codon_start=1
/db_xref="PID:g60379"
/db_xref="SWISS-PROT:Q01053"
/translation="MQEMTQWLNIRASCLGALAAPTFIWCIFIKSLFTFSIFSL
QISFVWVLGAHVTILCYCIITFKEMSYIIEALGACLFNMLTFEISHLWAPLY
TLPFVFLVCLSLVPIYIDIVLCPETIKYKYLGEFLNMLLYVMVAMRWLYSV
IWCPLVLGMEVFAKNEHPVENILITCKPFTAKNKKYKTEVMKLVNHN
LGAVLLAVFASVSVVQRLDIFVGMQNYLFLFVSMCCCMFSLPSNAICVVLST
FVCTVGNVSVALTCCNKSEQL"
complement(81056..82162)
/notes="ORF 59: BELF4; sim. to EBV BMRF1"
/codon_start=1
/db_xref="PID:g60380"
/db_xref="SWISS-PROT:Q01052"
/translation="MADTOTHSYGVGFKPCVLOEHTKIYNHIKAKVKGDIQIEGL
TSSPSRIWISGEGILSKINNVEFNCHMNPENVSFENISPGGNTFLYREL
EGCNVATLMSFNRSGKPEFTKAKLEYCDKYSTVTRHTSTIIPAGLFFVNVNSTV
LSRVSLPTAAMLQWLREHKAHTKVSVMNSTLAVLITVEDSTKVDLHIVDDMS
LCNNSKDGQVQVOTKVNITLALSAALGCKIPGVFPCIKFYESEVELEVCGTPYK
LGSWDTQVTLFCFPNLAEDTQAQVSDYFIPSPSACLADPADTQEEPTLNSCQD
SVSENPALPLKRRKREKDKKSKIKIAFNPLI"
complement(82273..83190)
/notes="ORF 60: BELF3"
/codon_start=1
/product="ribonucleotide reductase, small subunit"
/db_xref="PID:g60381"
/db_xref="SWISS-PROT:Q01038"
/translation="MSVKKYLTCBHVGLFELTKETWKNRWPSPVPLQDVGCCVDM
LNERDEFKFLFTFGMAKLNINIEDLSFDSDHSDISHIYAQOMAMENIHGVIA
NLNMLFKNNITVGYACIMDSALQEKRLWNLGRVTEASDKAEKILFLVLEGIF
FISFSEIGLFRVGMGICLANDIARDMLHTSAAALLYNTMTKSSRSEDWY
KLFRVAEVEFKFAAGYGVSLVNVNHEIRQLQATADRILESINLNPYGSUPPENC
PLAYTSKTSVNEFERDNDYCTGLNDL"
complement(83196..85499)
/notes="ORF 61: BELF2"
/codon_start=1
/product="ribonucleotide reductase, large subunit"
/db_xref="PID:g60382"
/db_xref="SWISS-PROT:Q01037"
/translation="MQSEITISNIDMLKVSAGWDRANEISGRLFHKLMDMSESTFI
SYMFLPGLLEPHILEFTQYQDEIDVELEYRASDYDFNCLRGLLPKAKRYDVTY
LPRTFMNGKLESIPIHFARIAAYCAWNCIMCEPLKDTLYVQVRDWNVETQMDQIF
KYFYKIVISCCATPMVRSGVAGENLSSCFIAPTLOTEKSTISSIFGELAPILLA
SRSGVDVTKFSFGKNTHSCKLINAQVEFFDNKSVRPVSVATYEVHWCQIHEFL
SAKLPNRNCNSIFQGVCPVPSFFKMYESDPNGLMTLPDQAPNLTLYGLEFEE
YLRVSEKIKQSVTLKSLMFLINTLIKTSPIYVTSKNAKHHWYETQGEAINCNS
LCAEIVQQKQFTSTNLANVCLPKCLNSNFYTSKTAQFDSEKLEYAVQAAVFI
NACILSPSTTSATVQGRSMGICGLADVFSEMGYGLDLESECLDRIDRETMYY
TAVKTSSEICSVSGQOPFAGFRSKLAHGVFWATWDAMPQVRPMQWHLQDNKKF
GVNSFOIALMPTAGTSQLTGYDSFPYFANMSSKVSNEEMKKNITELKNKKPD
LCTVRYIGDVSMPEDVSTRYKHELTAFDPCPEAQMRASIRAPYVDQOSQSLFTI
ENVOQAUKLMDLLGLFGLRGLTIMYICRVKTKTLQLLECLKLDHETTKDAQIVLA
DIARELPDSHKTEDACPLDQSECAQ"
complement(85499..86491)
/notes="ORF 62: BELF1; sim. to other HV"
/codon_start=1
/db_xref="PID:g60383"
/db_xref="SWISS-PROT:Q01051"

```

CDS

CDS

```

/translation="MPKVNENSFKKIEKDKVLKLLPPRPKHIEISEGQKFAKSIROALS
KYVTSVNHPIDEFNSQMLDVPFROPLVADFLIOAKTTHOKPERIGSFLTEKOEDSGSS
IDILFESVSEKISGLDAEPAPVHRISFVWYAGNCDNCTVPNLAELLEGKHFQFL
TPVGPLVENTISTFVTRIKSVVKGELNNKSPPENKULLPSDFIDUDEPNIDTLAP
KHLUSFYACVTPVLHNVPYMALTFKSNKSIKDLFCHLKVIYSDIIRTYIAMOND
FYINMFGFALCKLSSSETVPRRGNYHFRGSSLTVEVPDFVADPGPMWMTIL"
86498..89197
/notes="ORF 63: EERF1; sim. to other HV"
/codon_start=1
/db_xref="PID:g60384"
/db_xref="SWISS-PROT:Q01055"
/translation="MSVKPLTEFLKNLEAASNELLTKVKMLMELNLYLSIEQLESAQN
ITDFNLILKQTSOYSTFIHOHFLYLLKSLTSTFSTLVNLDLESIKOYMLNLCNDVAT
TIHATSSNENLNQTVINHIKQYITSNATFTGLSEPIVNVNISTFRCVEEVHVCFQ
CIWHFFPQAQIPOPNGALEKWLITQHFLNLDITAFNSLKDQATYILTHEKHLLFVP
LSSEYSLTLPKAKNALNIYTSFTNTITKSNVPVAFSAKELTDATPELFFYDFI
IEALYHEHSYNVQNIIEQFISKNTQFMDTLCNTIQKCSNKSTSSIRIKHELLES
CGLTEECRLQTSVLISNSVGTSSNKGYETFISLSQLVLSDFYKCLFYFRSPTS
IGHSKITELNTVSAISETLSHANKSMLANLSPFIPKAPSKILLETHTHSPLYL
MKSAFSIAKWTNYTLDATSPHTQHTFNHAPVISOSEVQKICEMLQGLTBYDS
RIVNSHULFAEEFITHIITLPTALONKQKRALQRLWLIVFASDAKGLYRIRP
LGLYFQIIEIFHDSNAAAAILNVDLYNEIQLOIYQVYTPYTFPIQIOELFSIKY
KPOSTELSKSIOKFAETECVODLPFQIQTNNMNTYVHIENYININIQGOSPAR
LDTKALHAKAIOGLKESWTTISQSKELQAYIQALILEIEKISQSIAKVS
NPNEIKANNFLQCFKKNYIANLITNSHSENLRIYFRIFEPELIPITTYOKILNFN
TETDDPQDFLSDSLPNTAPKSKELTSEDFNRLFEFANPVPETAPASSIKHYS
TFNTPQVNNKTYETHTYIADSPAELQFTHLTSAILDALSK"
89200..96609
/notes="ORF 64: EERF2; sim. to other HV"
/codon_start=1
/db_xref="PID:g60385"
/db_xref="SWISS-PROT:Q01056"
/translation="MDHPLFKLNLEGIASTHQADEKYGQYAGSQCLSNCFVLVS
YQSTDEPTVSLHGLEILYKAGIDFLRRSGQLGNQYAOHLHPIGVIAGPKWACFI
YQNTSEMGLGHGESPINEPFFVASLKLSSKNYNTVOYFLAICHKSGMILKOKKIF
LFDPSGCLPVNSPAHFVSTNSVNDAYEILSPNVQYIGSLYFVPRKYEIGHSHYIMN
HYRVINEKLGHPNIDLTQEGLEIEISPPNPKPTSTQPKPTPRTPKAPTKAPKT
PRKPKTESTIPYDKSKPKIPKTSKKSKVLTQDTALTPOKHTIBEHRELPLPI
TETVEDNTFNHPVERTPGDLSLGSINTSKREDDEDODDNTSKLEDODDWDID
IPIPELDTTETHSDQETIYMGIDENHDWSDDDDIDTIDISFIDJNLNLTSLDNI
PKNTPRIDKTSNQPIKEGKALSHDIRILKNVLEHGLTSSISISKSKSLQFV
ILWGEKSLIPTRDLTKLEIIEAIALTKLNTDFNNVITKLNKMLKUKSE
SVDSYKLSALLNNIILKIQITDEIEKLTSTVTSSELGDFSVYCTKSEETIMAA
IKNKEIKSTRKQKLTHTENYFQSLVAMETFPQIPTRVIEIQPSKKAQKHEKSK
LVEOKLIDRANVLTLHTMKQKDTDISPADFTYLNKQIQLLOLTQVTDNLID
KKFTSNVQQLSYIGWVAELSHSQWNEPKADVPKILDDIKKEIQOVTOKNEE
TSLKILADVQTLLENKQSDLSIPIQHYITKAGTLVGERENQKFSKNTYVKLST
SEEFKTLIDSTLLENVQLOQIEIDILSNQYIQHSETIKQAFDKSNTIINNLIQ
INQOYTYTQPLMIAVREFLSEAKFRESNTICEITSLVSLGSLSKSTVEALKDA
LKSIDTLKEKLTAVDRPLKRELVNVRKLOKLTLLQEQDFDNKMEVDSFVPTPSR
DKVTFIQNASPKAKOYAKKALQIOAMEIDVDPEVIEDNIKANGAKNQKIQSAF
QDLNFSILIPDDMLSLAKETRPAKSTFTVIGPILLKVEVEVLESVKNAKAKLSLL
PNGVPTPPKDFDWHIYESNVNHLKTLNPKVSTVAHNHIEJLSUALNSKTLPE
AVGTSLEQHAAKFSCHFLEATWHDQVDRTRKIDYIEDLRNDTKKHTVAQIQOS
PNRFLESEDQIEINSLPKFRDLSLENESRLASQKNEFQMLENTVKAELQVATQES
DIISNMEASLCLPLAPAYTLAIPITPDPLKVVENTIODKRLNTEPYOITIECLN
WLNTAKTILSICPKSQKQVLDQSDINTHLNTOQFYNLEKANTDITDLSVQNAL
STLDRKVGQKATVDSQKQKQKAMLDNKSQAQTLASLDLIMGTALTYSVSTHAL
GELLQKADPLQKDIESLUSSTNIDLLSRVTLEIHFIFKRGFLSYEEGQKVFQRYPL
TQNTRPSQPTIEINLLRALFVLKNDKASAWIWTETPLVDSNKLAVVPPNKQRYPLT
CSQYKLLKLEQLDPLSLKSVLSDNRPLAGTAQARLGIDSTVLLARAPEDTQKHAER
LTAYKNSISHTONEFAMTIVCHMIKITMNDYFQNFENINTVTVYVNHKTLQIOLIT
MWPLLIKASLQQSQSQEATSLQTLPLFLKITDLDLENNIYNPASHCSALLFPQ
KWSINQSIWNEHPFSLAICKNSRARITFLALAFKIIDPTILNQLMNTSLNPANTS
TYSVSLNHLNHLVATEFQNVNPTFLEPNPSLAYAQGTQGTNIIGTSYVQKSPPTIS
VTAPEIAGLALIFQVPVKLFTDKTVPVLSPELGDMLVISELLCTQTEPTEKMEIA
PKSSLNKNQYVSPPELEVFESRQASQHLHSSSNKNNIVATIDYSTTELNAVY
VPEKLPKQOSFCFIPKIDSLOWNNFTTFPLVEMPSNLELHAKVETFFNTVLS
TMNVFTHLTPQEEHDSISSKSPTFKIEHDINTNSVNNHNNLNNTNNSTYHQY
KDLVQPLADKLSYEPKDLQNLASTEPQIEDIFSELISKETDNTAKAPLYPQKQK
NKKFSPVHTKHKHTNSIIIEENTYKVPQNTCQHSDLHKDNTWTPQKISNAPCFIP
NHKVPYIKPSQEKLANNTDNLSPKKPQLIANNNNIFKQSKQKHQVTOIS
KPKFIFNQDSNNPIKQPHNPPQPKPTPDQQLTSLNDIISDQTTKLNLVKPII

```

CDS

CDS



```
CDS
VIPNNYALNOVKLSNPISIKTPYITLKDIQSNKTYLDESPTITPILEHLDIEPI
VSISLEKRVDETKEIILEFIKTKQNIKIITNLLIHQIMKTIYLV"
complement(96613..97032)
/note="ORF 65; ECLF7; sim. to EBV BFRF3"
/codon_start=1
/db_xref="PID:g60386"
/db_xref="SWISS-PROT:Q01047"
/translation="MHLRVDPVVGKLESDSHLVKRLIILPQNMTPEEYTLM
KRNLYVFLIAQNFQOYIETHGIRKKKHIEGLKANLPKLSALSUGLSYSLT
TPSSUSTPSTSPGLSISGPSTTIDSKKKPKAK"
complement(97025..98332)
/note="ORF 66; ECLF6; sim. to other Hv"
/codon_start=1
/db_xref="PID:g60387"
/db_xref="SWISS-PROT:Q01046"
/translation="MDTIOERYMDNLYKFGSGCDNFKVERTLNSLLQPLQIDSSDFI
KEVYGGVINEHCLPSPVFLDRCTSISELSFWCGIVMDTRRTQVHKFKIKLQTC
LFRAIVVVVFPKRDLPKPKFLENNWYKIINPFYKAIVTFLMLNLNLSIKHPLI
QFQSLPMDLSILRKKNLFCSTLPLSPAPSORRENENLVHSDDDQSHFALMA
ALKQOQAVPGCNPLDMKVLCSFMSQNKYAIIPMONIEKTFDVLVKILGYNIL
SSVGLPILCKIKDKIINKTYSQHIIVCIEGHCLNFGRGSKNLNFPPTTHVFCRD
QKQFTICGTSGRYICSGCSQSPKFPMPVEANIRAVIANNAACMAQCAQQQDDV
VPCLMGCGCIFKRVTVQSLYLTLSKIESLCCVKCGVIYNYA"
complement(98245..98952)
/note="ORF 67; ECLF5; sim. to other Hv"
/codon_start=1
/db_xref="PID:g60388"
/db_xref="SWISS-PROT:Q01045"
/translation="MNSFRLVVELCDIVNLYLQCPGVQIDVDRCASGPHVFTKGCTEA
ICTVKLSHLYNTEFVYKFAWKLESVKYFSPCFIISNGLATLAKCFSPRPNNV
HFGHLNLDSDVLTKNSTSVILSQDDFVKFTNLVFSKDLDFHSMVYFRLIEHRQ
ALQFLVYKPSKRVNLSLVAKTASQNFILDPFRSEETRVCIKPTLSKKNWTI
ILSLVAVAILIKKRWEL"
99198..100508
/note="ORF 68; ECRF1; sim. to other Hv"
/codon_start=1
/db_xref="PID:g60389"
/db_xref="SWISS-PROT:Q01040"
/translation="MIPQKSTYKHKTSIECLLNYGFMPTPETALDNLALVHTYA
ALTSTCKICQITLYSKNTSPAVSFYEDYSLCLTCLYAPITWTSLTMAADFIEI
IKTHPTSDNFYAPQSLAIDQLHFYIHRCFKVLSSNDILSSLSQFLKTTFLQG
KLGTGIPQCFCKFAWKIDCCNNTSHDLPSNLSSVFCADQLKPLNLLPILDIWS
ASDLFKNVNSQPFYFIEDICQCPGLSPSLGTQKNNTSICPLCECIASHP
NAIDTQTLKTIINCENNVKLDRIISFILSNDLDFIQDPIKATVIONCISOEIHK
HFFCDPCALANKITNLIKPIDPNLLKLCARLATGEHLSKNYLDCEYLETAL
IFKCSQCKVGTTFLEIIRELDLLSKKNIPTVKAFOTSQIA"
100511..101296
/note="ORF 69; ECRF2; sim. to other Hv"
/codon_start=1
/db_xref="PID:g60390"
/db_xref="SWISS-PROT:Q01041"
/translation="MSSRSVKRSKSTKSRHHHPYVKLDKMFSAISSKKELGTDFLR
EMDAPICTSKILLPLDLNLSIPGRCVLSLSPGSHNMFQCEKTESKNGSGDVSO
NHDIYSLVFLYKVDVKVKKHAYLSLSHSMENLKSFQPELLYAVVYKRGHN
VFPFPEKDCSLCLFKCTKQTHIGESCLRMNDMLNPISIDYIKDVMYAMFTQC
FAIORNISIAEDTICESVSTLDCDELREEIVKGINALQIKDI"
complement(101533..102417)
/note="ORF 70; ECLF4"
/codon_start=1
/product="thymidylate synthase"
/db_xref="PID:g60391"
/db_xref="SWISS-PROT:P06854"
/translation="MSTHEQHGHEQYLSQVQHILNYGSKFNDRTGTGTLISFQTSQ
RFSLENEFPLLTKRVFWRGVQELLWIRGSTDSKLSAGVHIWDANGSRSLDKL
GFYDRDGLDTPGVQWVGFQVGYKGVGRDYKGEVDQLQIDITIKNTDRLML
CANNYDIPKMYLPCHVLSQFYCDGKLSQQLQFORSADMGVFPFNATSLTLCMI
AHVNLVGFETHITIGDAHYVDHIDALKMQLTTRPFPPTLRFARNVSCIDDFKADD
IILENYPHPILIKHMAV"
complement(104685..105189)
/note="ORF 71; ECLF3"
/codon_start=1
/db_xref="PID:g60392"
/db_xref="SWISS-PROT:Q01044"

CDS
BASE COUNT 37678 a 20687 c 18269 g 36296 t
ORIGIN

Query Match 3.3%; Score 79.4; DB 17; Length 112930;
```

```
/translation="MDLTKTVLHITDSEDEMYCLIFLINGSICIPRSCNAVNLSDLI
ETLSKSTWDICLMOCLVLRKIGLLLNLFQVTKAVKOSFFQPOLETHVITLVNPN
NLTAKDERKRLCIFLDQFPFRNVAAPSVILCVSNMLCEMHVLECLICQLKCLKQIGR
SDIAKTV"
complement(105190..105954)
/note="ORF 72; ECLF2"
/codon_start=1
/product="cyclin homologue"
/db_xref="PID:g60393"
/db_xref="SWISS-PROT:Q01043"
/translation="MADSPNLRNRAKIDSTTMKDPVLNMLKRLLELLPKFTSLWFIQ
TEVTVDNRTILLTMHLLCESELDKSVPLSVSLDRYLCKKQTKKTLQIGRACV
LIGSKIRVVKPMTVSKLYSLSCDFTNLEINQEKDILEAKLDEATVLAIDLPLIC
NALKIPEDLPOLYEAASTICKALIQNIALLSPLGLICAGGLLTITETDNTNCRPWT
CYLEDSSILNFTSTNVTVKQVSEAFSLYDLEIL"
complement(106010..107233)
/note="ORF 73; ECLF1"
/codon_start=1
/db_xref="PID:g60394"
/db_xref="SWISS-PROT:Q01042"
/translation="MAPRRKAKRRHRLRSECKDKCKHVQYVSPRRKRRKLKPOG
DDIINTHQOAAALTEQORREVEBERERBREREGEGEGEGEGEGEGEGEGEGEAE
EKAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE
EAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE
EAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE
RSRLPKGSHDKVPKKYQARNKFFSOAAPSVIDLSPKSWCVWVFWGPTDALYRLSR
SLSPFGVSSGIGTPEPKGPHATGPWYFITYVTCRTQTAKEVIKAQKYEKKYPRSAK
LKASLGKSKSLPIE"
107732..108697
/note="ORF 74; ECRF3"
/codon_start=1
/product="G protein-coupled receptor homologue"
/db_xref="PID:g60395"
/db_xref="SWISS-PROT:Q01035"
/translation="MEVKLDFSEDFSNYSYSGDIYGVDPVCVWNLISESALAF
IYLMFLCAIAGNSLVRLFLKIRAQOAFDYLMMGFCLNSLFLAGILMLRLRMFEI
FMNTELCLEAFNLNLSYSPFLLVLSVLCILIFLCATLWVKTKLIGVFLCCSF
VLACFGALPHVMTSYEPSSICEEDVLTQRLTKLNTFTWTFAGFLFTVTCYS
MSCYKLFKTKLSKRAEVYITMTTLLFVFCIPYIMESIDTLRVGVIEETCAKRS
ATVYGIOCTMLLYCYMLPFAMFGLSRQMAAMCKTICH"
complement(108782..112681)
/note="ORF 75; EILF1; sim. to ORF 03 and EBV BNRF1"
/codon_start=1
/db_xref="PID:g60396"
/db_xref="SWISS-PROT:P11282"
/translation="MAANRHGHLPLPEGLAAPTHTVYIYAKCDFSPSEERCVRNPT
GRPGLTLMRORSNVEHVLTVSYGNLHQRLLTSALEQORACKTKLILLLAPQI
DYNPLTLDLHNLGQRAVIFSYGNLHQRLLTSALEQORACKTKLILLLAPQI
FTSKAVQYVNEEDILTAALIGENNYINTETSLATRLVDALSWAEYKPLDVTCH
AVMKPPREGVQPMCLVASPSVDYFNFKYMITPCTATSLHMGHPYPRVSGOIFHVH
SSTRMGQCCDYLHSSLLPNGNTAFACGYVTSTGSGSIPYVNEOILHAASAO
GRSLNLAQVYVSGFLPLPCRSSEVPNNVITHVSQLRASERDNLNCRVACQFVAAV
GAYDTPSGDKSPILYRDSIDSLRAIATKLFQMCETPCVSSYQREFGSCSTLHLH
LALVSPKMTVHISKLPETKALRSVPVLEEDTVCAFVGYFFNCFSLQFLVDIDDK
VKTPSGGIIHFTDILKAGNLKAPVYILGRCTNDIGHCVNDLYHPRDLVLDQAT
SSMLITVOPHAAVSATLQPOBHEDESIDWMEGTSTSTNISLHSAVASKNIIR
RDLRCGNGLIAQPGIGSDAPVSDYAIICDSSMPARLENDASQIKISKQEAQARAY
AQIHKWCTEKLFLNTISAKYIALGEQAYKLSRNPVGVKVAIAIAINIMPGDCLV
EDILTVAAHNKQETALYRVLFACKMRELNVNLSITDASDRSDPTQIDTAANT
VVVTASARVTSIERITPALKAEALHVSCLSKELTSGSVFENSTAFSSHLPLDT
KSLDMFYAVKHLISKNLVAGHSDIGDLITTAEMCFASTFVTVNLPSPALPALMY
LVSETPGALLVPKREHLSVTLLSERGLTWAVGTNNVNNLSIYDNGHTLLTESIN
ILNSKWSVAESPEPCEPHIESMYNDYGNMMDLKHLEDCLTHKPLQLYTCSHPV
SVAVLTPGCDPDPVATLQAFANFGLSPISTEFELQGNLNAFSLAVSGSSAEHEE
GYTRIALYITLQCDLAKNLEKFFQRDLTSLCCGELGTACACQVQVGDTHPRGDD
ISSNPETLEPNNAKSHESLWNLNHFVPTTKSIILQALRGITFQDGLQWGLGRLY
KHDAQEYIMQNGTIAMSYHSAKNIPYIYAMHYPRNPSSNGSVAGICSKNGRSHALLV
EPALSFHTWQWQHPKPLVTPSWALMYQCMFLWCVE"
37678 a 20687 c 18269 g 36296 t
```





```

polyA_signal
5761..5766
/note="EDRF3 ORF"
complement(5793..5798)
/note="EDLF1 ORF"
complement(5797..5399)
/note="first methionine codon in the EDLF1 ORF"
/codon_start=1
/db_xref="PID:g330984"
/translation="MQWLHSCWPGFKYRQKVVPLDTSSTSPDRWKIEIEIAQPGV
FVGDILQNSDSASRQAYLLAVQLNNTIDYLRKDEASVPESCKSVQVQITKRSV
RNIWNTMLSMVAGVGTIDDAALKTLDRKAGESIALIEMLAIVMDSDSKAWAKE
INNIISAEHEKQILVNSEVPLECETLAEEKTTTPAVSI"
6378..8885
/note="first methionine codon in the EDRF4 ORF"
/codon_start=1
/function="required for DNA replication"
/db_xref="PID:g330985"
/translation="MNEATEAWPKFKVLFATDGDSEIITDILGTDTNFIYSVLHN
CYIYEVKIVILCLPAKPGGDKLEVFOLHIDTELAIPFLVTKPLKANDLHKY
IDFKARKNFKPILDIISNPKSPKTHNSDIKSIWFAKFNLSRLKLTSSSPYW
MITTGSVEVPELLAIFVFFQHNCTINTIFHLSLFEKKLGTSIAITTEELGGV
CSTDYLTAPAFINCHIKILARDESQAIDSITDRGLMSQNDLVHYIYISFF
CITHLPDLIGYAPQDCTSFYVWAGQCKNVHNLVINTPHNISDGLDLDAID
OCLNIDIFIKYSHLNSDNIHFVPETELVAQLDENFRKMLTYNNKLYIYITHK
STFVNDLKDQVNESOKVPYVRCFLNKTYEVIQNDILKNVSTDVLMQENWYM
LKDEITSNISYKETFMTSLTRDQLKISRHEYFNRLPVFNLDLHLHITSEHEI
DEIYNLCTLSLILETLOGLGPDVIDTHHVFIFKSTCEKNWINDNKKELFCYCTKK
LGFRIITLPAGVILGNSPVIISFVNLNRTIKDKKLLAMYPLINETDPDFDGIYH
KRCVRIPTHYVNSGSLRELKLFVCHPHVDNKLQIYVMSDFDINLLYHSHNPKV
KOLKAVYDIADTNENFLOKAQALPQTNHNAVERIESASHMSITDWAFAWPLFE
LIKLYLSEKVSQFHVSPAAGTGNLIKITISLGNFSCLEFKHRLKTSVRIFLSLH
TPDNCVTLTASQCFASKNSCKTAHMSVRPIDK"
8881..8886
/note="EDRF4 ORF"
complement(9074..9079)
/note="BELF5 ORF"
<9082..10320
/note="5' end of coding region for EJRF1 undetermined."
/codon_start=1
/product="transactivator"
/db_xref="PID:g330986"
/translation="VLGSSDDDFDSSSDSSDEESTSPQIMKSDVTWASPPSPPEP
SPDVASVSNLKRQRSPRTWEHQSLRSVYRSPRFRGPRPRISSNSTSRKSTS
WADRYREAAQRPRSPRKPYSHPNGPLRNGPPRAPLLKFLDISILPKSGEPTN
LPVSLPCOEAKTNDKYVLAQAMHDVPISKKOLANLPPKFLPILLSVRYTPN
YYWYMRKETASNLCTAAAFDELSCGOQLKNDIFSENGKDIILDSALLS
OLVHKIMLPLFCMLQMTPQDHIQVQCYLIASNNRILDVAVRLQTSVIKSPVILL
AYAVCLPAAICTKNETOLYSHCMRLKREYRGDVMNLLHESLTQHLNKKCPSTCAYT
TRAVGTANTGTGLEFLPTQ"
10425..10430
/note="EJRF1 ORF"
complement(10679..11752)
/note="first methionine codon in the BELF5 ORF"
/codon_start=1
/db_xref="PID:g330987"
/translation="MIQEMWTQWLVNIRASCLGAATAFPINCFIKSLFTSFISL
QISIFYVLAGAHVILCYITFTKWSYIEALGACFLTMLTFTFISHLWAPLY
TLFPVFLNCLSLWPIYDVLVCPITYKVELGLFNAMLLYVYVAVNRMYLSV
IFMCPFVLGLMGVFNKFNHFPVENLITCKPFTAKNKYKGTGVNKKLVYHN
LGAVFLFLAVPASYVSVQRDLDFVQWNYLFLFVLMCCMSLPSNALCVLET
FAVTLIIVHLVDKPLSIVIGGLLVIVILLMQAIGCOIETIRTKLAGDVGPKLCL
FLCTVCNIVSVLALPCCNKSQEL"
complement(11147..11152)
/note="BELF4 ORF"
complement(11749..12855)
/note="first methionine codon in the BELF4 ORF"
/codon_start=1
/function="required for DNA replication"
/product="DNA-binding protein"
/db_xref="PID:g330988"
/translation="WADQTTHTFSYGVGKPCVLQEHKIYNIHAKVKDGLIQIEGL
TSSPSLRWSSIGELISFKINNVSFVFNCHMPENVSFNRNISPQGNFLYREL
FGCNVKTATLMFNSRSGKPFPIKALEYCDKRVSTRHTTIPAGILPVENVNSTVS
LSRVLSPKTAAMKWLREHKAHKTIVSVMNSTLAVLILTVEDSTKTVLDLHIVDDMS
LCNNSKDWGVQNDVQTKVNITALSALGLCKIPGVPCIKKEYSEVLEVCGTVPK
LGSWDTSLQVILFCTPNLAEDTQAASDYFFISPSACLTADPADTQEEPTLSSNCQD
SSVENSPLPKRRKEDINKIKIAFNPLI"
complement(11774..11779)
/note="BELF2 ORF"
complement(11774..11779)
/note="BELF3 ORF"
complement(12966..13883)
/standard_name="ribonucleotide reductase beta subunit"
/EC_number="1.17.4.1"
/note="first methionine codon in the BELF3 ORF"
/codon_start=1
/product="ribonucleotide reductase small subunit"
/db_xref="PID:g330989"
/translation="MDSVKKLYTCDHVGFLTKETKMRNMFPSQVPLQGDVCCVDM
LNERDLFEYKFLFTFLGMAEKLNINIELLSQFSDSHDISHYAEQAMENIHGKYVA
NILNMLFKNITEVSYACDINMSALQELKRLWGRVTEASDKAKILLLVEGIF
FISSEFISGLFRVGTIMGICLANDYIARDMLHT"SAALLYNTHTKSSERSEDIY
KLFRAVEVEFRKIAKGVSVLNVHEIRQLQATADRILLESINLNIYISLPPENC
PLAYTSSTKSVNFFERDSDYTGTLTNDL"
complement(13889..16192)
/standard_name="ribonucleotide reductase alpha subunit"
/EC_number="1.17.4.1"
/note="first methionine codon in the BELF2 ORF"
/codon_start=1
/product="ribonucleotide reductase large subunit"
/db_xref="PID:g330990"
/translation="NSQETIISNLDMLKVSAGDREANEISGRFLPKHLMDSSTETI
SOYMSLFGLEPHLEFIQYEQEIDEVLEYRASYDFMCLRNCGILPAKRFYDYY
LPRTEMNGKYESIPHFARIAACANWCMCEPLKDTLVYVQKRWNVVEIKDMQIF
KIFYKVISQVCCATPVMSRAGVAGENLSSCFIAPTLDTERKSTIISFGEIAPLLA
SRSGVDVDTKFSFGCKNIHCLINAQVEFFNDKSVRSPVATYIEVHQCIHEFL
SAKLPEPDRCSIFQGVCSPLFFKMSDNGMLYLPQDAPNLRLYGLFEFEE
YLRVSEKYYKOSVTLKSLMFLINTIITKTSYVITSKAMKHHWYEQGAENCS
LCAEIVQOPKQFTSCNLANVCLPKCLANSNFPYTCSTAQDESKLEYAVOAAVLI
NACILSPSTSATVQQRERSMGICHLGADVSEMGYGLDLESCLDRDLFETMYI
TAVYSSEICSVGKQGFAGFRKSLAHGVFWATWDMQPRVPMKOWHLODNIKKF
GYNSQFIALMPTAGTSQTYDTSFYFANNKSVKNEEIMFNTFLKNVKPQD
LCTVRFYGGDVSMPEDVSTRYKHTAFDYCPAQMRRASTRAPVDQSQSILFLT
ENVOSAKYLLKDLLGLLFRGLKTIWYCRVKTKTLQLECLKLDHEHTKKAQIVLA
DIARELPSHKTEDACPLDQSECIACQ"
complement(15105..15110)
/note="BELF1 ORF"
complement(16192..17184)
/note="first methionine codon in the BELF1 ORF"
/codon_start=1
/product="capsid protein"
/db_xref="PID:g330991"
/translation="MPKVNSFKKIEKDVLKLLPPRHKIFISEQKFAKSIQALS
KYTVSNHPIDFFSNQMLDVPFRQPLYADFLQAQTIHQKEPIGSPFLTFKQSDGSS
IDIIFSPVSFYKISGLDAEFAPSVHRISFVWAGADNCIEDTVNLAELIEGKFHOFI
TVPGLVENITSTFVTRIKSVVKGVLNNKSPENVKLLLPDLDFIDDEPNIDTLP
KHLVFIYACVTYIPVHLNVPVNFALFIKNSIRDLFCHLKYIISDIITKTIAMQD
FYINKMSFALCKLGSSETVPRGNVHFRGSSLPTEVDFVADPGPWWTIL"
17191..19890
/note="first methionine codon in the EERF1 ORF"
/codon_start=1
/db_xref="PID:g330992"
/translation="MSVKPLTELTKNLEAASNELLKTKVLMDEMLNLYSIEOLSAQN
ITDFNLKQTSQYSTIHFHFLYLLKSTFTSTINYDLESIKQYMNINNVCDVAT
TTHATSSNNLNQTVINHKOYITSNATFTGLSEPIVPNNVISYFRCVEEVVHFQF
CYWHEPFOAQIPQIPNGALEKLLTQHKFELNDYAFNLSKDAQTYLLTHEKHLFVP
LSSSEYSLTLPKNAQNLNITSFTNTITKSNVPLAFSAKELDTATBELFDIFI
IEALYHEHSYNQNIIEQISKTFQMDICNTIQIKNSKLSYSSEIRHRELLES
CGLTECCRLTQTSVLIINVSFTSNWGYETFISLSQVLFSPFYKFLTFYSPTS
IGHSKITELTNTVSAIESHTSHANKFSWKLNLSFFLPKAPSKILLTYTHISPYL
MKSAFSIWAKTWNTVLTADSPHTOTHTFKHAPVISQSEVQKVCENLOLQTEYDS
RIVNSHLFAEETHTHIIPTLAILQNKVKNALEQLRWLIVFASDEAKGLYRRRP
LGLLQFQIETIEPHDSNAAAALLNVDLYNEQQIQLQIYVPTIYPIKFIELFSLKY
KQSFELSKIOKFAETECVKDILFPFLGTNMCNTYIHTNININIQGSPAR
LDTKALTHAIKAIQIETESWTTISQSYKELOATYIQLATILETEKISHOIAIKVS
NPNEFIKNNTFLQCFKCKYNTIANLITNSHFNLTFRQIFPELIPITVTKILNFN

```

```
DETDDPQFLDLSQPLSYSHYNAPKKSELTSEDFNRLLEFANPNVETAPSSIKLHYSD
TFTNPQVNIWKKTYTEHTYTIADSPAELOFTHLTSAIDAELSK"
19893. .27302
/notes="first methionine codon in the EERF2 ORF"
/codon_start=1
/product="tegument protein"
/db_xref="PID:g330993"
/translation="MDIHPFKKLNLEGIASHTOADERYQAGYQSCLSNCVMELVSS
YNDTEPVLSHGNEILKYGAKIDFILRRSGOLGHNOYAOHLHPIGVYIAGPKWACFI
YQSTEMGLGHPESHPINEPFAVSLKSLSKYNTVVOYELAICNSKSGMLIKKAKIF
IEDPHSPVPNSPAHVESTNDAIEYLSPPNVQYTGSLFVVPKYEYIGHSHYIMN
HYRVINEKHPNIDLSQEGELIEISPPNPKPTQKPKPTKPTKPTKPTKPTKPTKPT
PRKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPT
TEVEDNTFNHPVERTPTGDSLLSGNSTTKRDDLEDODDNDTLDISFQIDNLNLTSDNI
IPIPEVLDTEHTHSDQETIYMGIDENHDMYSDDDDIDTLDISFQIDNLNLTSDNI
KNNTPFRIDKTSNQPIKEGKALHSIDRIKLVLEGLITSSSISISKCKSLQFV
ILWGEKLSIPRDLKTLIKTLITIEIAELTKLNTDNGFNNTIKLNMKLUKSE
SVDSYKLSALLNIIILKIQITDIEIKLITSTVFTSGNDFSVYCKKESETIMAA
IKNKKISTRKQELHTEENYFQSVLIAMETFPPLTRVIEIQPSKKAQOLHEKSK
LVEOKLIDANNVLDLHTMKQKTDIISPAPDFTVLKNIQSTLQQLQTKQVTDLNID
KFTISNVQOOLSYIGWEAELSHSQWNEPKADVPILKILDDIKKEIQOVTQKNEE
TSLKILADQLLENKQSDSLIPILQHYIYKAGLVGERENQRFESKNTVQKLS
SEEFKLIDITLLENQOQIEISDILOSNOYIHOSETIKOAFEDKNTIINNTLQ
INQOYITVQPMIYAVKRFLESEAKFRESNTICEITLVLGSLSSKSTVYEAALDA
LKSIDTLKEKLTAVDRPLKRLXNVIRKLOKQLTLBQOEDFNKMEVDSVPPTPSR
DKVTEIONAPSKAKQAKVAKLQKIQAMEIDVOPEVIEDNIKANGQAKQKIQSAF
QDLNFSILIPDWLSLAKETRPKSTLFTVIGPILLLKFEVLESVKNKAKLSLL
PNGVPVPPKFDWIHYESNVHLKTLNLPKSVTAHNIHGHESLLSQALNSKTLPE
AVGTSLQEAHAFSCMKFLEATWHQVDRTRKIDIEYIEDLRNDRKHHVAPQIQS
PNRFLSPEDIOEINSLKFLRDLLENESRVLASOKNEFQMLENTVKAELQYKATQE
DIISNMEASINSLAPAYLAIPTDPLKYVENIIOKRLNLTPEYQITIECLN
WLTNACKTLISCPKSQORLVLDOSINTHLNIQOYNLEKNTANTDDLSLQNAL
STLDLRQVGGKAVDSQSKQKQKAMDNISKAQTLASLDIWMGALTLSVSTAHL
GELIQADQPKLIESUSNTDLSRVTELHFKFRGLFSYEEQOKEVFPQYPL
TQNTRPQSTPINLLALFVLKNDASAWITETPLVDSNKLAYEPTTEPKTIEA
CSQYKLLLEASHQSDSKSVILSDNRPLAGIQAARLDSTVLLARAPDPQKHAEEV
LTAYKNSIVSTHNEFMATVICHMIKIIMDVFQPNFINTPVLYVNHKLLQILIT
MWPLILKASLCQSQEQEATSLQTLKPLFKLITDITLLENNINPASHCSALLFFPQ
KWSKINQISNWEHSPFAJCNKRSARITELAFKIDITDILNQLMNTSNPANTSD
STYSLLHLNVAFTEDKNPSTLEPNPSLAYAYGTQGTGNIIGTSYVQKSPPIIS
VTAPEIAGALIFOVVKLVFTDKTPLVSLSPDELGMVSELDCGTTEPKTIEA
PKSISLNLNKOYVSPHELEVESROASWLOHLSNKNKNIVATIDYSTFLNAVY
VPEKLPKQESFEPKIDSLQWPNNTTETPLVEMPSNELHVAKYTEPKNVLS
TMNVFPPIHLTQEHQDSKSPKIEHDYNTSVYNNHNNINLTNNSTTHQY
KDLVPLADLSDKDLQNLASTEPQIEDIFSELSKETDNTAKAPLAPKQKOP
TKKFLSPVHTKHTKSNISFEENTVQVQNTQIQLSHLKDNTPROQISNAPCFIP
NHKVPVITKPSQEKLKANTVHTDLSPPKPOILANNNIFKQSDKOHQYTOIS
KPKLFINQDSNPKIKOPHNPPQLIKPTDQOQLTSLNDILSSQDTTKLNAVORPIL
VIPNNVALNQVKLSNLPISIKTPYILKDIQSNKSLYDESPTIIFLEHLDTEPI
VSIYLEKRQVDETFFIILEFIKTKNQNIKTNTLLIHIQIMKIKTLYL"
21293. .21298
/notes="EERF1 ORF"
complement(27216. .27221)
/notes="ECLF7 ORF"
complement(27306. .27725)
/notes="first methionine codon in the ECLF7 ORF"
/codon_start=1
/db_xref="PID:g330994"
/translation="MHRLRVDPVQVQKLESDSHFELVKRLLEILPQNNMPEEYTLM
KRNVLVLIAGYNQDYQETIYTHGIRKXKHIGLKRANLAPSAISASLSGLLSLST
TPSLUSSTPSTSPGLSIGPSTDTIDSKKKPKAK"
complement(27520. .27525)
/notes="ECLF6 ORF"
complement(27718. .29025)
/notes="first methionine codon in the ECLF6 ORF"
/codon_start=1
/db_xref="PID:g330995"
/translation="MDTIQERYMNYLFSGCGDNFKVETLNSLLQPLQKSDQFI
KFVYGGCWINHCHLPSPYFLDRCSTISELFSWCGIWMVDRRTQVHKFKLIKTKTOC
LFRAIVVVVVPKCRLDKPKFLENVWVKYINWPYKALVTEMLNLSIKHPLI
QFQSCLPWDLISILRKNKLFCSLTNPLSVAPSPQRRENLENFVHSDFDQDSHEPALMA
ALQKQKQVPCGNPLDAMIKVLCFNSMIQNKIYAIIPMDNIETNFDLVLKILGNIL
SSVFLGPLICKIKDKIKNTYSQHIIVICIEGHCILNFGRGSKNLNFPPTHFVYCRD
OKIKOFTTCGTSGRYCYGCGSQRKPKPMVEANTIRAVIANNAACMAQCASQOQFDV
VPLCGMCGSCIFKRVTVQSLYLTISKIESLCCVKCGSVIYNA"
28032. .28037
/notes="EERF2 ORF"
complement(28733. .28738)
/notes="ECLF5 ORF"
complement(28938. .29645)
/notes="first methionine in the ECLF5 ORF"
/codon_start=1
/db_xref="PID:g330996"
/translation="MNSRKLVELCDIVNLYLCQPVQIDVRCASGPHVFTKGTGEA
ICTHGLGVNIIEFYKFWAHLKSVKPYSPCFIISNGLATLTKCFLSRPRNVN
HFGHVLNIDSDVLYLNTKNTSQQDDVFVKTNLVSFKDLDFHSMVVFRTYLIEHRQ
ALQFVKNVPRSSKRVNSILSVAKTASQNFILPPRRSETRVCIKPWTLSKKNWTI
ILSLAVVAIILKWL"
29891. .31201
/notes="first methionine codon in the ECLF1 ORF"
/codon_start=1
/db_xref="PID:g330997"
/translation="MIIPQKSTLYKHKTSIECLLNYSPMPTPETALDNALVHTYA
ALTSSTCKICQTLISLKNTPAVSFYEDYSLLCLTCLYAPITWTSLMTADFTETI
IKTHPTSDTSNFIAPQSLAIDIOHFYIHRCFKVLSSNDILSLTKFTLFTFQOG
KITGIPQCFKTAIKNNDTCNNTSHDLNLSVFCADQLKPLNLLPILDIQWS
ASDLFNKNVNSEQFPFTYPEDIDICQPCLLSPSLGTQKNTTICICLCEIASHP
NAIDLTQIKYTIINCIENNVLLDRISFILSNDLDFIQDPTIKTVIONCSQETIHK
HFFCPQCALNKKSTNLFKLIPDNLKVLICARLATGEHLKSNYLCCEYLETIAL
IFKQSTQKVGKTFLEIIRLELDSKKNIPTVRAFQTSQIYA"
31204. .31989
/notes="first methionine codon in the ECRF2 ORF"
/codon_start=1
/db_xref="PID:g330998"
/translation="MSSRSVKRSKTSRRHHPYVKLTKMFFSAISSKELGTDFLR
ENDAPICSTILLPLDNLNISPGRCIYLSFGHNSMFEQCEKTESKNKGSDVSQ
NHDLVSITLVFYKNDVKVYKFAFVLSLSHSMENLKSTQPELLYAVVYVKEAGN
VPIFFEDCLCSICITFKCQTLHIGESCLRMNLNPNYKISIDYIKDVYAMTFTQC
FAIONISTAEDTICESVSTLCTDELREIIVKGINALQIKDI"
31483. .31488
/notes="ECLF1 ORF"
complement(32145. .32150)
/notes="ECLF4 ORF"
complement(32226. .33110)
/EC_number="2.1.1.45"
/notes="first methionine codon in the ECLF4 ORF"
/codon_start=1
/product="thymidylate synthase"
/db_xref="PID:g330999"
/translation="KSTHTEQGHQYLSQVQHILNYGSKFNDRTGTGLSTFGTQS
RFSLENEPPLTTKRVFWRGVYVELLWFRIGSDSKLSAAGVHIWDANGSRSLDKL
GEYDRDEGLGPVYQWRHFGAEGYKGVGRDYGEGVDQLKIDITKTNPTDRRML
CWMYSDIPKMYLPCHVLSYFYCDGKLSQYQSRADMGVGFNFIASYSLLTGM
AHVTNLVGEFHTIGDAHIYVDHIDALKMQLTRTPRPPTLRFARNVSCIDDFKADD
IILENVNPHPIIKMHMV"
33164. .33169
/notes="EERF2 ORF"
complement(35245. .35250)
/notes="ECLF1 ORF"
complement(35245. .35250)
/notes="ECLF3 ORF"
complement(35245. .35250)
/notes="ECLF2 ORF"
complement(35379. .35882)
/notes="first methionine codon in the ECLF3 ORF"
/codon_start=1
/db_xref="PID:g331000"
/translation="MDLKTIVLHITDFTDEMYCLLFLINGCIPRSCNAVNSDLII
ETLSKSTQMDICLMQCYVIRKIGLLNLFQVTKKAVQSFQPELTYHVLTVNWN
NLTAKDEKRLCFLDQFFPRVAAPSVILCVFSNMLCEMHVLECLQLKCKLQICR
SDLAKYV"
complement(35883. .36647)
/notes="the first methionine codon in the ECLF2 ORF that is
homologous to D-type cyclins; putative"
/codon_start=1
/db_xref="PID:g331001"
```



```

Db 1605 CTGAAGAAGCTGAAGAAGAAGCTGAAGAAGCTGAAGAAGCTGAAGAAGCTGAAGAAGCTG 1546
QY 782 AAGAAGTAAGCATTTTCCCTGCTGAAGAACAGCAGGAGTACCACCACTACTTAAAGCT 841
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1545 AAGAAGAAGCTGAAGAAGCTGAAGAAGGAGGAAGAGGAGGACCAAGTACTCCAGCTT 1486
QY 842 T 842
Db 1485 T 1485

RESULT 15
AC005505/c
LOCUS
DEFINITION
ACCESSION
NID
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
***
*** WARNING: Phase 1 High Throughput Genome Sequence ***
***
* This sequence is unfinished. It consists of 17 contigs for
* which the order is not known; their order in this record is
* arbitrary. In some cases, the exact lengths of the gaps
* between the contigs are also unknown; these gaps are
* as runs of N as a convenience only. When sequencing is complete,
* the sequence data presented in this record will be replaced
* by a single finished sequence with the same accession number.
*
* 1 6598: contig of 6598 bp in length
* 6399 6648: gap of unknown length
* 6649 15165: contig of 8517 bp in length
* 15166 15215: gap of unknown length
* 15216 34693: contig of 19478 bp in length
* 34694 34743: gap of unknown length
* 34744 41106: contig of 6363 bp in length
* 41107 41157: gap of unknown length
* 41158 64588: contig of 23432 bp in length
* 64589 71511: contig of 6873 bp in length
* 71512 71561: gap of unknown length
* 71562 74477: contig of 2915 bp in length
* 74478 74527: gap of unknown length
* 74528 89168: contig of 14642 bp in length
* 89169 89218: gap of unknown length
* 89219 93019: contig of 3801 bp in length
* 93020 93069: gap of unknown length
* 93070 122669: contig of 29600 bp in length
* 122670 122719: gap of unknown length
* 122720 139586: contig of 16867 bp in length
* 139587 139636: gap of unknown length
* 139637 153447: contig of 15811 bp in length
* 153448 155497: gap of unknown length
* 155498 167106: contig of 11609 bp in length
* 167107 167156: gap of unknown length
* 167157 173414: contig of 6258 bp in length
* 173415 173465: gap of unknown length
* 173466 179108: contig of 5644 bp in length
* 179109 179158: gap of unknown length

```

Search completed: May 14, 1999, 01:44:28  
Job time: 1688 sec